THE EFFECT OF PLOT CO-REGISTRATION ERROR ON THE STRENGTH OF REGRESSION BETWEEN LIDAR CANOPY METRICS AND TOTAL STANDING VOLUME IN A *PINUS RADIATA* FOREST

A dissertation submitted in partial fulfilment of the requirements for

the Bachelor of Forestry Science Degree with Honours.

Benjamin Thomas Slui

New Zealand School of Forestry

University of Canterbury

2014

ABSTRACT

Background: The objective of this study was to verify the effect that plot locational errors, termed plot co-registration errors, have on the strength of regression between LiDAR canopy metrics and the measured total standing volume (TSV) of plots in a *Pinus radiata* forest.

Methods: A 737 hectare plantation of mature *Pinus radiata* located in Northern Hawkes Bay was selected for the study. This forest had been measured in a pre-harvest inventory and had aerial LiDAR assessment. The location of plots was verified using a survey-grade GPS. Least square linear regression models were developed to predict TSV from LiDAR canopy metrics for a sample of 204 plots. The regression strength, accuracy and bias was compared for models developed using either the actual (verified) or the incorrect (intended) locations for these plots. The change to the LiDAR canopy metrics after the plot co-registration errors was also established.

Results: The plot co-registration error in the sample ranged from 0.7 m to 70.3 m, with an average linear spatial error of 10.6 m. The plot co-registration errors substantially reduced the strength of regression between LiDAR canopy metrics and TSV, as the model developed from the actual plot locations had an R^2 of 44%, while the model developed from the incorrect plot locations had an R^2 of 19%. The greatest reductions in model strength occurred when there was less than a 60% overlap between the plots defined by correct and incorrect locations. Higher plot co-registration errors also caused significant changes to the height and density LiDAR canopy metrics that were used in the regression models. The lower percentile elevation LiDAR metrics were more sensitive to plot co-registration errors, compared to higher percentile metrics.

Conclusion: Plot co-registration errors have a significant effect on the strength of regressions formed between TSV and LiDAR canopy metrics. This indicates that accurate measurements of plot locations are necessary to fully utilise LiDAR for inventory purposes in forests of *Pinus radiata*.

Keywords: *LiDAR, Plot co-registration error, Pinus radiata, radiata pine, LiDAR-based forest inventory.*

TABLE OF CONTENTS

ABSTRACT i

GLOSSARY iii

1. INTRODUCTION 1

2. BACKGROUND AND LITERATURE REVIEW 2
   - 2.1. The Application of LiDAR to the Forest Inventory: 2
   - 2.2. Regression Strength and the LiDAR-Based Forest Inventory: 3
   - 2.3. Plot Co-registration Errors and the LiDAR-Based Forest Inventory: 4

3. PROBLEM STATEMENT 7

4. RESEARCH QUESTIONS 7

5. METHODS 8
   - 5.1. Stand Information for the Trial Site: 8
   - 5.2. The Establishment of the Plot Co-registration Error in the Trial Forest: 9
   - 5.3. The LiDAR Dataset: 10
   - 5.4. Adjustment of the Pre-Harvest Inventory Dataset: 11
   - 5.5. Data Analysis: 12

6. RESULTS 14
   - 6.1. Plot Co-registration Error: 14
   - 6.2. The Effect of the Plot Co-registration Error on Regression Strength: 16
   - 6.3. Model Strength after Stratification of the Plot Co-registration Errors: 20
   - 6.4. The Effect of Plot Co-registration Error on Model LiDAR Canopy Metrics: 22
   - 6.5. The Effect of Plot Co-registration Error on the Percentile Elevation Metrics 27

7. DISCUSSION 30
   - 7.1. Explanation of the Results and Implications: 30
   - 7.2. Critical Evaluation of the Findings: 32
   - 7.3. Future Research: 34
   - 7.4. Study Limitations: 34

8. CONCLUSIONS 37

ACKNOWLEDGEMENTS 38

REFERENCES 39

APPENDICES 41

GLOSSARY

Actual Plot Position: The survey grade GPS verified location for a plot in the forest.

H_{20} Metric: The height of the lowest 20% of all LiDAR point returns above the ground level within a plot.

H_{95} Metric: The height of the lowest 95% of all LiDAR point returns above the ground level within a plot.

Intended Plot Position: The location that the forest owner had intended that the pre-harvest inventory crew would establish a plot centre in the forest.

LiDAR: Light Detection and Ranging. A form of active remote sensing that measures the location and three-dimensional shape of objects within an environment.

LiDAR Canopy Metrics: Statistical descriptions of the height, variability and density of the point returns in the LiDAR point cloud.

LiDAR Percentile Elevation Metrics: The LiDAR canopy metrics which quantify the height of point returns in the point cloud at threshold (percentile) levels.

LiDAR Point Cloud: A database of geographic co-ordinates for the locations where the LiDAR light pulses had reflected from objects in the environment, such as the ground terrain or forest canopy.

PCR>3 Metric: The proportion of canopy returns that are above 3 metres in a plot, relative to the total number of the first point returns in each plot.

Plot Co-registration Error: An error in the location of field plots, which causes the plot locations used in LiDAR analyses to be spatially asynchronous with their true positions.

1. INTRODUCTION

Light Detection and Ranging (LiDAR) technology is a form of active remote sensing that measures the location and three-dimensional shape of objects within an environment. This technology is becoming an important tool for use in the pre-harvest inventory because LiDAR data can be applied to improve the understanding of the forest resource (Wulder et al., 2008). This is through the development of plot level relationships between LiDAR data and measured stand inventory variables, which are used to predict stand inventory variables across an entire forest estate (Wulder et al., 2008, Adams et al., 2011).

It has been recommended that in order to achieve accurate and valuable results from the LiDAR based forest inventory it is imperative to accurately define the location of field plots (White et al., 2013, Adams et al., 2011). This is because the accurate identification of plot centres would ensure that correct LiDAR canopy data is regressed against the stand inventory variables that were measured in the forest inventory. Spatial errors in the location of the pre-harvest inventory plots, termed plot co-registration errors, could result in the use of incorrect LiDAR data in the regression models that are used to predict stand variables. This could reduce the accuracy of the stand variable estimates produced from the LiDAR-based regression models.

The objective of this study was to verify the effect of plot co-registration errors on an operational LiDAR-based forest inventory for a mature plantation forest of *Pinus radiata*. This was achieved by determining the level of plot co-registration error which resulted from using the intended, rather than survey grade GPS verified locations, for plot centres in a LiDAR analysis. Then the effect of this plot co-registration error on the strength of regression between LiDAR canopy data and the measured total standing volume of plots was assessed. The changes to the individual plot-level canopy metrics, which were used as independent variables in the regression models, were also quantified.

This research was undertaken to provide a greater understanding for the influence of plot co-registration errors on the operational LiDAR-based forest inventory in the New Zealand forestry setting. This will indicate the value of accurately identifying the location of field plots for the LiDAR-based forest inventory.

2. BACKGROUND AND LITERATURE REVIEW

A recent focal point of the remote sensing research for the LiDAR-based forest inventory has been in understanding the factors which influence the strength of regression between LiDAR data and measured stand inventory variables. This is because the strength of regression between these variables will underpin the value of LiDAR data to the forest inventory (White et al., 2013). Importantly, this research has identified that the strength of LiDAR-based regressions can be influenced by plot locational errors (Gobakken and Naesett, 2009; Frazer et al., 2011; Marshall 2012). The purpose of this literature review is to provide a background for the LiDAR-based forest inventory and to review the research which has demonstrated that plot location error, termed plot co-registration error, is an important consideration to this process.

2.1. The Application of LiDAR to the Forest Inventory:

The acquisition of LiDAR data for the forest inventory typically involves using a specialised aircraft to fly over a forest estate and to measure the time taken for light pulses (which are usually calibrated to the near-infrared spectrum) to be emitted and reflected back to the aircraft (Adams et al., 2011). This process produces a LiDAR point cloud, which is a database of geographic co-ordinates for the locations where the LiDAR light pulses had reflected from objects in the environment, such as the ground terrain or forest canopy (Wulder et al., 2008). The point cloud quantifies the three-dimensional structure of an environment, including the form of the ground surface (Krogstad & Schiess, 2004); as well as the height, structure and variability of above ground features, including the forest canopy (Popescu, 2007).

In order to contribute value to the pre-harvest forest inventory, the LiDAR point cloud data requires extensive analysis, to convert its geographic point return information into an estimate of stand inventory variables. The analysis procedure required to achieve this involves initially filtering the LiDAR point cloud data-points to identify the underlying bare earth (the digital terrain model) and to standardise the heights of canopy returns above this ground level (Adams et al., 2013). The standardised canopy point cloud is then clipped to the boundaries of known plot locations, to produce separate point clouds for each plot that was measured in the pre-harvest inventory (White et al., 2013). The plot-level point clouds are then processed into LiDAR canopy metrics, which are statistical descriptions of the height, variability and density of the above-ground point returns in each plot (White et al., 2013). The plot level canopy metrics are then regressed against the stand inventory variables which had been measured from the corresponding plots (White et al., 2013). Typically, linear regression or nearest neighbour modelling are the statistical procedures that are used to regress the stand variables to the LiDAR canopy data (Eskelson et al., 2009, Dash et al., 2013). These models are used to predict stand variables across a forest estate.

2.2. <u>Regression Strength and the LiDAR-Based Forest Inventory:</u>

Importantly, the value of LiDAR to the forest inventory will be determined by the strength of the regression models that are used to predict stand variables from LiDAR data (Wulder et al., 2008, Adams et al., 2011). This is because a higher strength of regression will result in a greater accuracy when the LiDAR-based models are used to make 'wall to wall' estimates of stand variables across forest estates (White et al., 2013).

In New Zealand research has demonstrated that moderate to strong relationships between LiDAR data and stand variables can be produced in *Pinus radiata* forests. From which, it is evident that the estimation of tree height is one of the most suitable uses for LiDAR technology. This is because a national model has been developed to predict mean top tree height in plots from a LiDAR canopy height metric, which had a high model R^2 of 95% and a residual mean square error of 1.91 metres (Watt & Watt, 2013). The strength of regression between LiDAR metrics and the total standing volume (TSV) of plots can also be strong, as a national survey has produced a model to predict plot TSV from LiDAR canopy metrics, which had a model an R^2 of 83% (Watt & Watt, 2013). However, the strength of the LiDAR based regressions can vary. For example, a study in a Bay of Plenty forest of New Zealand recorded a weak relationship (R^2 = 42%) for a linear model that used a LiDAR height metric to predict total standing volume (Marshall et al., 2012). The implication of the weaker model strength is that it would reduce the ability to make accurate predictions of stand variables from LiDAR data.

Given the importance of attaining high model strength, it will be beneficial to understand and control the factors that influence the strength of relationship between LiDAR data and the bio-physical features of the trees. This will maximise the value that is gained from the

LiDAR based forest inventory. Importantly, the research which has investigated the drivers of model strength have identified that plot co-registration error, whereby the plot locations used in LiDAR analyses are spatially asynchronous with their true positions, is a contributing factor to the strength of regression formed between LiDAR data and stand inventory variables (Gobakken and Naesett, 2009, Frazer et al., 2011). The causes and effects of plot co-registration errors on the LiDAR-based inventory are reviewed below.

2.3. <u>Plot Co-registration Errors and the LiDAR-Based Forest Inventory:</u>

Plot co-registration errors can result from a variety of factors during the manual pre-harvest inventory which limit the ability to accurately define plot locations. A primary cause for this error can arise from the inaccuracy of the GPS (global positioning system) units that are used to measure the locations of plots during the pre-harvest inventory (Johnson and Barton, 2004). This is because the geographic positioning accuracy of GPS units can be limited by clock-time discrepancies, atmospheric effects and multi-path error (Johnson and Barton, 2004). Alternatively, plot co-registration errors may also arise from the management decision to not measure the locations of plots with a GPS and to instead use an assumed location for each plot in the LiDAR-based forest inventory, which may be incorrect. These causes could result in a high occurrence of plot co-registration errors in typical forest inventories, and hence the effect of these errors could be considerable.

The effect of plot co-registration errors on the LiDAR-based forest inventory has been a focal point of three different studies. Gobakken and Naesett (2009) initially demonstrated the effect of plot co-registration error in a mixed species forest of Norway spruce (*Picea abies* (L.) Karst) and Scots pine (*Pinus sylvestris* L.), that had been measured with aerial LiDAR data. The plot co-registration error in this study was simulated by randomly shifting plots by 1 to 20 metres from their true locations (which were determined using a differentially corrected GPS). Their results indicated that the development of LiDAR metrics which quantified the heights of point returns within plots were strongly affected by the plot location errors. Furthermore, the mean error in the prediction of Lorey's mean height, basal area and standing volume from the LiDAR based regression models also increased with higher levels of co-registration error. It was concluded that plot co-registration errors of at least 5 metres were most limiting and that larger plot sizes were less affected by the plot co-registration errors, compared to smaller sized plots.

Frazer et al (2011) also analysed the effect of plot co-registration error on the development of regression models which predicted forest biomass from LiDAR metrics. This study was stimulation based, where an artificial forest was designed through computer modelling and a synthetic LiDAR point cloud was generated over the canopy. The forest developed was a mixed aged Douglas fir (*Pseudotsuga menziesii*) stand which had substantial variation in its canopy structure. The plots in the artificial forest were shifted using the Monte Carlo procedure by 1 to 5 metres, as this was argued to best reflect the error of GPS units. Their results indicated that the mean absolute difference in the LiDAR canopy metrics which described the height, variation and density of plot point clouds, increased with higher levels of co-registration error. The greatest changes to these LiDAR canopy metrics occurred in the most heterogeneous areas of their forest. Furthermore, the co-efficient of determination (R^2) for the linear models that were developed to predict total above ground biomass from LiDAR canopy metrics decreased with the increasing co-registration errors. As an example, the R^2 of their models reduced from 92% to 87% after the 5 m plot co-registration error (for plots with a radius of 10 metres). This provides evidence to demonstrate that small locational errors can adversely impact the strength of LiDAR-based regressions.

The effect of plot locational errors in New Zealand forests of *Pinus radiata* was tested in an un-published study by Marshall (2012). This assessed how shifting plots from their true locations (as verified using a survey grade GPS) by up to 25 metres using the Monte Carlo procedure, affected the strength of regression when predicting the total standing volume of plots. The study had a replicated design and formed 30 separate regressions from 124 plots at every one metre interval of co-registration error. The results indicated that on average the model R^2 of the regressions used to predict the total standing volume (TSV) from the 30th percentile elevation LiDAR canopy metric, only declined slightly after the plot location errors (as the mean model R^2 was around 50% for all levels of co-registration error surveyed). This indicates that plot locational errors only had a minor influence on the LiDAR-based forest inventory in the *Pinus radiata* plantation.

However, Marshall (2012) also demonstrated that the variation in the model R^2 between LiDAR data and TSV increased with higher plot co-registration errors. For example, the range in model R^2 values was between 45% and 55% for the 30 regressions that were produced after a 5 metre co-registration error, but this range increased to between 37% and 61% for the models which were affected by a 20 to 25 metre plot co-registration error. This is significant, as it indicates that higher plot co-registration errors caused both increases and decreases to the strength of regression between LiDAR data and TSV.

Overall, these studies have demonstrated that plot co-registration errors will have an effect on the regression produced from the LiDAR-based forest inventory. However, an important distinction is that there was a disparity in the effects recorded in each of these studies. This is because the average strength of regression was relatively less sensitive to co-registration error in the forests of *Pinus radiata*, as trialled by Marshall (2012), compared to the mixed age and species forests as trialled by Gobakken and Naesett (2009) and Frazer et al (2011). Furthermore, Marshall (2012) had also demonstrated that increasing locational errors can also result in increases to model strength. This result contrasts directly to the findings by Frazer et al. (2011), which indicated that plot location errors only reduced the strength of LiDAR-based regressions.

The difference in the findings of these studies could have occurred as a result of the differing canopy structure types in the forests trialled. As Gobakken and Naesett (2009) and Frazer et al (2011) had evaluated the effect of plot co-registration error in forests that are expected to have had a much higher variability in canopy structure, due to the variety of different species and age classes present within their respective forests. In contrast, the forests of *Pinus radiata* in the study reported by Marshall (2012) would be likely to have a more homogenous canopy structure, due to the similar genetics, silviculture and age of the trees within these plantations. This could have caused the differential responses in the strength of the LiDAR-based regression models to the plot co-registration errors. However, there is little evidence to currently support this hypothesis.

This review has established that plot co-registration errors are regarded as a limitation to the LiDAR-based forest inventory. However, it is apparent that there is an inconsistency in the recorded effect of plot locational errors on the strength of regression models that are formed between LiDAR data and stand variables. Furthermore, it is also evident that these studies have only used simulation-based analysis procedures to examine the effect of plot co-registration errors on the strength of regression models. Although this is an appropriate analysis technique, the effect of plot co-registration errors has not yet been evaluated in the context of an actual forest LiDAR-based forest inventory. This indicates that there is currently an incomplete understanding for the effect of plot co-registration errors on the LiDAR-based forest inventory.

3. PROBLEM STATEMENT

There is a need for further research to verify the effect of plot co-registration errors on the operational LiDAR-based forest inventory in *Pinus radiata* forests. This is because there is a disparity in the research literature which has assessed the influence of plot co-registration errors on the strength of the regression between LiDAR data and measured forest inventory variables. As the influence of plot co-registration errors on the strength of LiDAR-based regressions formed in *Pinus radiata* forests (Marshall, 2012) were inconsistent to the findings that were observed in two published studies (Gobakken and Naesett, 2009; Frazer et al., 2011). Furthermore, the effect of plot location errors has not yet been comprehensively evaluated in the context of an actual LiDAR-based forest inventory. This is because the previous studies have only used simulation analysis to determine the effect of plot location errors on LiDAR-based regressions. Rather, the effect of plot co-registration errors which would result from management decisions, such as using the intended positions for field plots (instead of their GPS verified locations) in the LiDAR-based forest inventory, has not yet been evaluated. Therefore, additional research is required to provide verification for the effect of actual plot co-registration errors on the strength of regression achieved between LiDAR data and measured stand variables in an operational *Pinus radiata* forest inventory.

4. RESEARCH QUESTIONS

1. Does the management decision to use the intended positions for plots, rather than survey grade GPS verified positions, cause a significant plot co-registration error?

2. Do plot co-registration errors significantly reduce the strength of regression between plot-level LiDAR canopy metrics and total standing volume in a mature forest of *Pinus radiata*?

3. Do plot co-registration errors significantly affect the plot-level LiDAR canopy metrics that are used to predict total standing volume?

5. METHODS

5.1. Stand Information for the Trial Site:

The study was conducted in Waipapa Forest, a 737 hectare plantation of *Pinus radiata* located 18 kilometres southwest of Wairoa, New Zealand (39.06S, 177.14E). The forest was composed of twenty individual stands which were; established in either 1986, 1987 and 1989, pruned to a height of between 4.5 to 6.9 metres and thinned to a final target crop stocking of between 250 to 350 stems per hectare. The topography of the forest varied from steep to rolling terrain, with an elevation range of between 0 and 250 metres above the sea level. Within the forest there were many canopy gaps caused by wind-throw, forest roads and intersecting power lines. There was very little understory vegetation in the forest.

A standard manual pre-harvest inventory was completed in the forest between July 2009 and February 2013. This was conducted by an independent forest inventory contracting crew. The locations of the field plots to be measured by the crew were established by the forest owner, which used a systematic, random sampling procedure to select one plot per hectare from each stand and excluded plots from canopy gaps greater than 0.1 hectares in size. These plot locations were provided to the inventory crew as GPS locatable waypoints and were also presented on a printed map, with 10 metre contours, at a scale of 1:1000. The field crew used a Garmin 60Cx recreational grade GPS to locate these field plots. However, the actual locations for the field plots were not measured with this GPS by the crew.

The plots established were circular and bound and had a size of either 400 m^2, 500 m^2 or 600 m^2. The field crew clearly marked the centre tree of each plot with spray paint to detail the plot number, with a circular band around the stem at the breast height level. All other trees in each plot were banded with their measurement number. Every plot tree had a diameter measurement at 1.4 metres in height and the pruned height, sweep and branch sizes were also documented for each stem. Up to seven trees per plot were measured for height with a vertex hypsometer. All inventory data was recorded and transferred to an Atlas Cruiser database.

The plot level inventory measurements for each stand were grown forward to the date of LiDAR acquisition (August 2013). This was achieved by using growth models and functions in Atlas Cruiser that had been developed specifically for Pan Pac Forest Products Hawkes Bay forest estate.

5.2. The Establishment of the Plot Co-registration Error in the Trial Forest:

The experimental design of this study involved comparing the strength of regression between LiDAR data and the total standing volume of plots; when using either the actual or incorrect locations for the measured pre-harvest inventory plots. The actual position for each plot was determined by using a survey grade GPS (the Trimble Geo 6000 XH) to measure the locations in the forest where each field plot was established. The intended positions of the field plots were used as the incorrect locations for each plot. These were the positions where the forest owner had intended that the manual pre-harvest inventory crew would establish each plot centre.

To achieve this, 204 field plots were selected for the study using a systematic cluster selection procedure. ArcGIS was used to randomly overlay a square grid of 580 metres by 580 metres above the entire forest; where all plots that were within 160 metres of each grid intersection were selected for the sample. A map of the plots selected for the study is provided in Appendix 1. The measurement of the actual locations for the sample of 204 plots was conducted between the 25th of November and the 14th of December, 2013. The Trimble Geo 6000 XH survey grade GPS was used to measure the actual locations of these plots. This was achieved by locating the centre of each plot in the forest by identifying the position of the marked centre tree and the centre peg. When the centre peg was not visible, the plot centre was established by determining the geometric centre of all marked plot trees. The Trimble GPS was programmed to derive a locational waypoint at each plot centre from 700 individual readings. The unit was raised 40 cm above the ground and any low level vegetation that may have interfered with the signal was removed before waypoint collection. The GPS waypoint data was post processed by Interpine Forestry Ltd.

The plot location error for the sample was established by calculating the linear distance between the actual and the intended locations of the 204 field plots. The proportion of common area overlap between the intended and actual plot boundaries was also derived. These calculations were determined using length and area measurement functions within ArcGIS. Figure 1 provides an indication of the spatial differences between the intended and actual positions of plots, from which the linear spatial error and the proportion of common overlap area, was established.

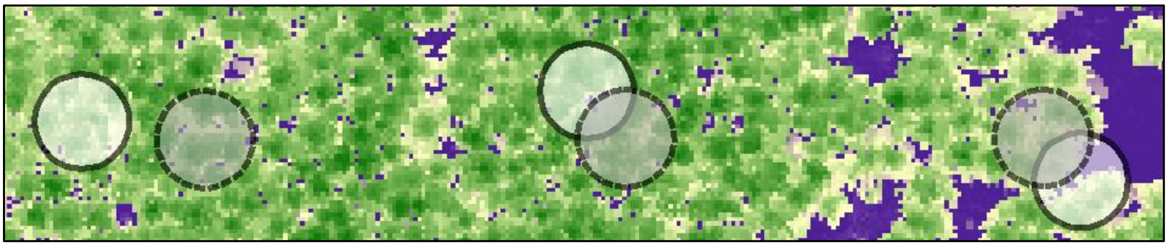

Figure 1: The difference between the intended location for plots (dotted circle) and the actual location of plots (full circle). Background raster = LiDAR surface layer (green colours indicate forest canopy).

5.3. The LiDAR Dataset:

The LiDAR dataset was acquired in August 2013 by NZ Aerial Mapping LTD using an Optech M200 LiDAR system and a CS8900 medium format digital camera. This produced a LiDAR point cloud for the entire forest area. The collection parameters are outlined in Table 1. The LiDAR dataset was supplied in the LAS format and was projected into the New Zealand Transverse Mercator (NZTM) co-ordinate system.

Table 1: Summary of LiDAR collection parameters from the acquisition of Waipapa Forest

LiDAR Attribute	Value
Flying Height (m)	1,200
Scan Angle (degrees)	20
Pulse Rate Frequency (kHz)	70
Mirror Scan Frequency (Hz)	33
Swath Side Overlap (%)	55
Minimum pulse density (points per square meter)	2
Average pulse density (points per square meter)	3.37

The point file produced from the LiDAR survey of Waipapa Forest was analysed using FUSION (US Forest Service, 2013) in January 2014. Initially, the ground returns were filtered from the point cloud to produce a digital terrain map (DTM), with a 1 square metre resolution. The DTM was used to standardise the heights of the canopy returns above the ground level. All of the point returns that were within 0.5 metres of the DTM were excluded from the analysis, to reduce the effect of understory vegetation on the canopy metrics produced. The ClipData function in FUSION was used to extract the standardised canopy data from within the boundaries of the actual and intended locations for each of the 204 plots in the forest. Figure 2 provides an example of four LiDAR point clouds that were clipped to the bounds of four plots. The intended and actual plot point clouds were then processed into statistical metrics, which defined the height, variability and density of LiDAR point returns from within the bounds of each plot location, using the CanopyMetrics function in FUSION. This analysis produced two datasets of canopy metrics (based from the actual and intended plot locations) for each of the 204 plots.

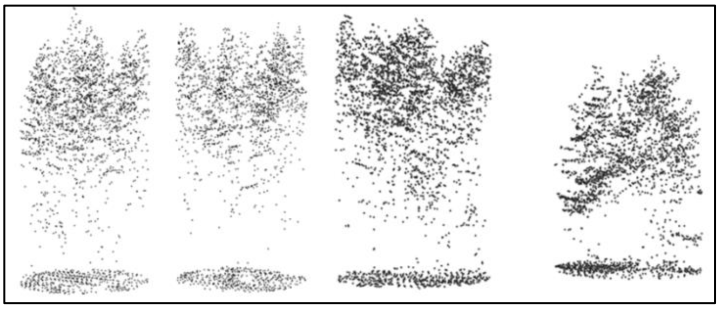

Figure 2: A comparison of LiDAR point clouds extracted from four plots in Waipapa forest.

5.4. <u>Adjustment of the Pre-Harvest Inventory Dataset:</u>

A preliminary study (conducting using the same dataset in January 2014) indicated that the mean top height (MTH) measurements that were determined from the manual pre-harvest inventory in Waipapa Forest would not be suitable for this study. The justification for this is outlined in Appendix 2. To overcome this limitation, the mean top height, *MTH*, for each plot was re-estimated using the following national LiDAR-based model, where H_{95} is the 95th percentile elevation LiDAR canopy metric (Watt & Watt, 2013);

$$MTH = 2.442 + 0.992 * H_{95} \quad (1)$$

This model is reported to be accurate and unbiased, with an R^2 of 95% and a root-mean-square error of 1.91 m (Watt & Watt, 2013). The 95th percentile LiDAR canopy metric (H_{95}) was extracted from the actual locations of each plot, as verified by the survey grade

GPS. Then using the estimates of MTH determined from Equation 1, the total standing volume (TSV) within each plot was then re-calculated using the national Peterson Equation (Kimberley & Beets, 2007) which is provided below, where G is the basal area measured from the manual pre-harvest inventory.

$$TSV = H_t\, G\, (0.942\, (H_t - 1.4)^{-1.161} + 0.317) \qquad (2)$$

5.5. <u>Data Analysis:</u>

The effect of the plot co-registration errors on the strength of linear regression models formed between LiDAR canopy metrics and total standing volume (TSV) of plots, was established by developing and comparing two multiple linear regression models. These models both predicted the measured TSV of plots, but were derived from LiDAR canopy metrics which were calculated from either the intended or actual plot locations.

100 different LiDAR canopy metrics were initially considered for both of the regression models. These canopy metrics described the height, density and variability of the point returns within the intended and actual plot locations. Stepwise selection was used to determine the inclusion of LiDAR canopy metrics as independent variables in each of the two (actual and intended plot location derived) regression models. The LiDAR metrics were selected successively for each regression model by a forwards selection process, where metrics were elected by their strength of correlation to plot TSV. Only statistically significant metrics were included (*p*-value <0.05) and metric selection ceased after the regression co-efficient of determination (R^2) did not increase by at least 5%. Furthermore, only one height metric was selected to mitigate the effects of colinearity. The model residuals were also plotted against each of the independent variables in the model to ensure that there was no bias (data not shown). Finally, a scaled power transformation was also applied to the LiDAR metrics, used as independent variables in both models, to reduce the model bias and ensure that model residuals were normality distributed.

The two linear models, which were developed from either the correct (actual) or incorrect (intended) plot locations for the entire sample of 204 plots, were then compared to assess the effect of the plot co-registration errors on the regression models. The co-efficient of determination (R^2) was used to examine model precision, while an analysis of the model residual values and their relationship to the model predicted values was used to determine the bias of the regression models.

The effect of increasing plot co-registration errors on regression strength was then assessed by stratifying the plots in the sample by their level of positional error and by separately developing regressions models to estimate TSV from LiDAR canopy metrics, for each of the stratified groups. The plots were sub-set into five stratified groups which either had low co-registration errors (common overlap areas of 80-100% or 60-80% between their actual or intended plot locations), or high co-registration errors (common overlap areas of 40-60% or 20-40% or 0-20% between their actual and intended plot locations). For each stratified group, two models of best fit were produced to predict plot TSV, by using LiDAR canopy metrics that were calculated from either the actual (correct) or intended (incorrect) plot locations. The model R^2 values for the each of regressions produced were then compared within each of the stratified groups, to evaluate the difference in model strength after co-registration errors, relative to the ‘no plot co-registration error’ reference point. Comparison between the stratified groups established the difference in model strength after low or high levels of plot co-registration error.

The effect of plot co-registration errors on the plot-level LiDAR canopy metrics was also examined. This was established by comparing the differences in each metric between the actual and intended plot positions and by comparing this change to the level of plot co-registration error. The metrics assessed included the LiDAR canopy metrics which were used as independent variables in the regression models, as well as the LiDAR percentile elevation metrics. The absolute differences in these LiDAR canopy metrics were plotted against the percentage of plot area overlap between the actual and intended plot locations. Tukey’s Honest Significant Difference Procedure was used to compare the absolute changes to these metrics within each of the five stratified groups. This established the changes to the LiDAR canopy metrics that occurred after increasing plot locational errors.

Finally, to establish the significance of the relationship between the plot co-registration errors and the changes to the LiDAR metrics, variable transformation was required for both the co-registration error (independent variable) and the change to the LiDAR canopy metrics (dependant variables), in order to produce an unbiased linear model. The transformed regressions are provided in Appendix 3. However, the model R^2 and *p*-values from these regressions for each LiDAR canopy metric are provided in the report.

6. RESULTS

6.1. Plot Co-registration Error:

This section presents the level of plot co-registration error between the intended and actual positions for the plots sampled in Waipapa Forest. This will establish the positional inaccuracy that occurred when using the intended positions for these plots instead of the survey grade GPS verified plot locations in the LiDAR-based forest inventory.

The frequency distribution for the linear distances between the intended and actual positions for the plots is presented in Figure 3. The distances ranged between 0.6 m and 70.3 m. The average plot co-registration error was 10.6 m with a 95% confidence interval of 1.2 m, indicating that there was statistically significant linear plot location error in the sample. The frequency distribution has a positive skew, as 82% of the intended plot locations were within 15 m of their actual locations, while only 5% of plots had distances of 30 m or more between their intended and actual positions.

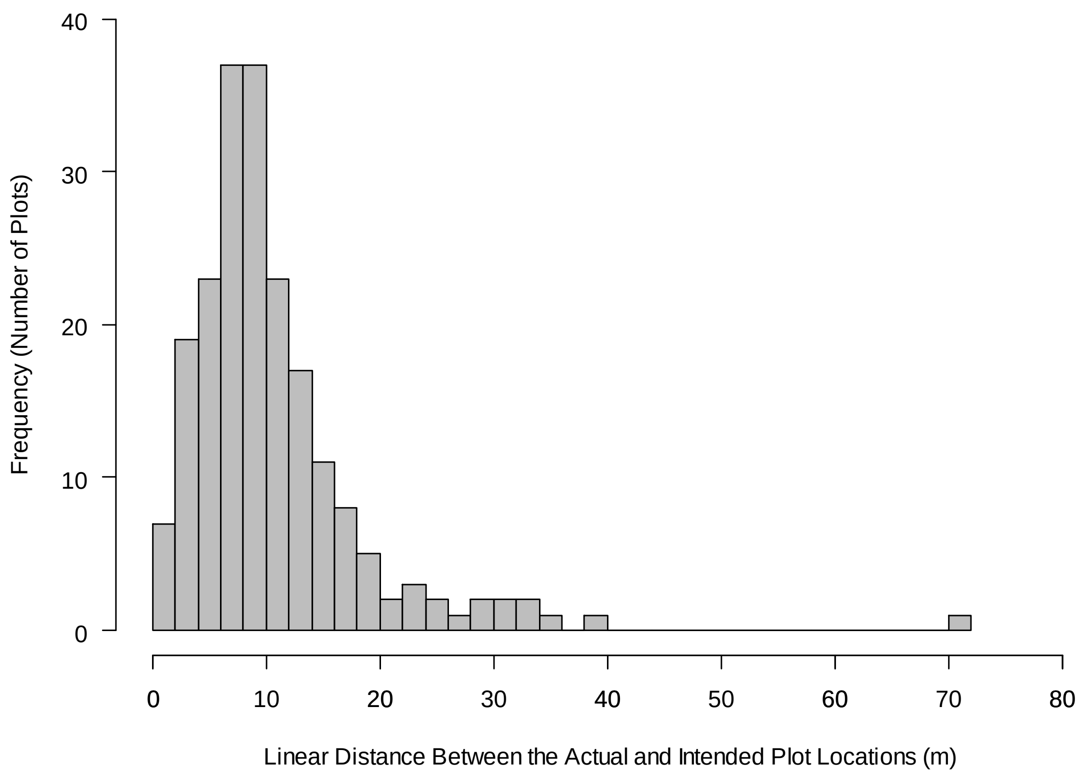

Figure 3: Frequency distribution of the linear distance between the intended and actual positions of plots for the sample of 204 plots.

The frequency distribution for the proportion of common overlap area between the boundaries of the intended and actual positions for each plot is presented in Figure 4. The lower proportions of plot overlap equate to greater plot co-registration errors, as this indicates that the intended and actual positions for a plot were further separated and shared a lower common area of overlap between their boundaries. The average plot overlap for the sample was 52%, with a 95% confidence interval of 3.4%. Only 6% of the plots sampled had 0% overlap (a complete separation) between their actual and intended positions.

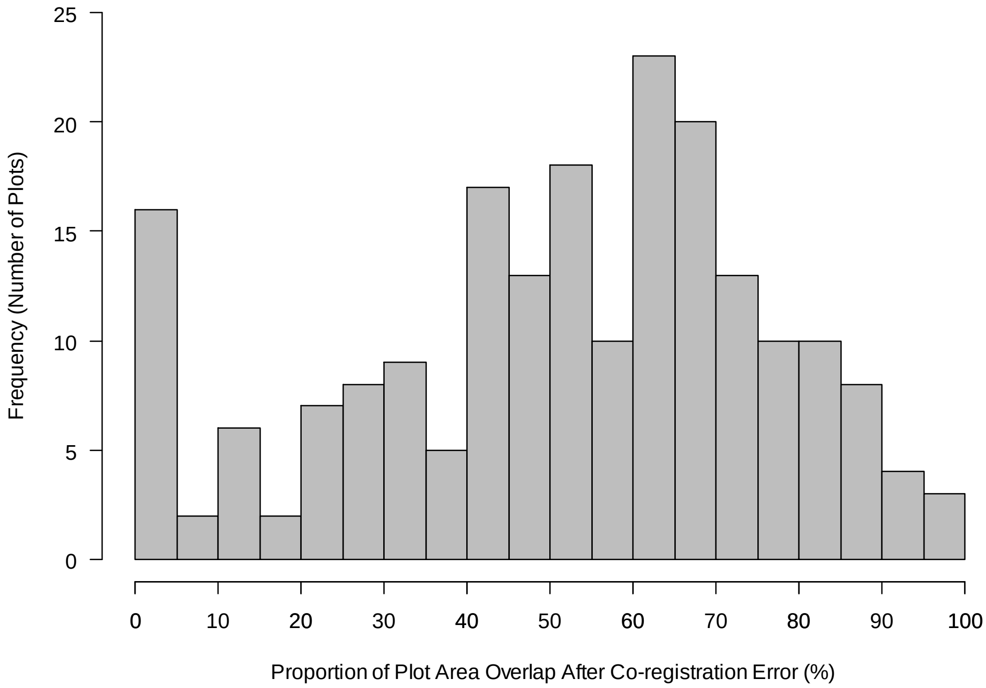

Figure 4: Frequency distribution for the overlap between the intended and actual plot boundaries for the sample of 204 plots.

Importantly, the average proportion of common overlap area between the intended and actual plot positions differed between the different plot sizes sampled in this study. This is because the average proportion of overlap area for plots of 400 m^2 in size was lowest at 41%, while the average overlap for the 500 m^2 plots was 57% and the average overlap between the 600 m^2 plots was highest at 63%. Tukey's Honest Significant Difference Test (Tukey's HSD) indicated that there was a significant difference in the percentage of overlap between the 400 m^2 and 500 m^2 plot sizes (p-value <0.001), as well as between the 400 m^2 and 600 m^2 plot sizes (p-value <0.001). This indicates that the larger plot sizes retained higher proportions of plot area overlap with their correct locations after the linear spatial errors, in comparison to the smaller sized plots.

For this study, the proportion of plot area overlap will be the measure used to define the level of co-registration error between the intended and actual plot positions. This is because this measure standardises the plot locational error (on a relative basis) for each of the differing plot sizes that were sampled in this study.

6.2. The Effect of the Plot Co-registration Error on Regression Strength:

This section demonstrates the effect of the plot co-registration errors on the relationship strength between LiDAR canopy metrics and the measured total standing volume (TSV) of plots. This is achieved by comparing two linear models which estimated the TSV of plots from LiDAR canopy metrics, which were calculated using either the actual or intended locations for the plots.

Importantly, both of the linear models were produced using the same two LiDAR canopy metrics. This included the 20th percentile elevation LiDAR metric (referred to as H_{20} hereafter) and the (All returns above 3.00 / Total first returns) LiDAR metric (referred to as PCR>3 hereafter). The H_{20} metric measures the height of the lowest 20% of LiDAR point returns above the ground level within a plot. The PCR>3 metric measures the proportion of canopy returns that are above 3 m in a plot, relative to the total number of the first point returns for each plot.

The optimal linear regression model between the measured total standing volume and the LiDAR canopy metrics extracted from the actual locations of field plots was:

$$TSV = 263.6 + 1.79x10^{-1}(H_{20}{}^{1.15/2.15}) + 1.7x10^{5}(PCR>3^{4.20/5.20}) \qquad (3)$$

This model required scaled power transformations with a lambda value of 2.15 for the H_{20} metric and a lambda value of 5.20 for the PCR>3 metric. The model had a significant intercept (p-value <0.001) as well as significant slope coefficients for the transformed H_{20} and PCR>3 metrics (both p-values <0.001). Overall, the model was statistically significant, with a p-value <0.001 and had an adjusted R^2 of 44%. The relationship between the LiDAR estimate of plot TSV and the actual TSV for each plot is presented in Figure 5. The residual standard error in the TSV estimates of the model was 150.4 m^3/ha.

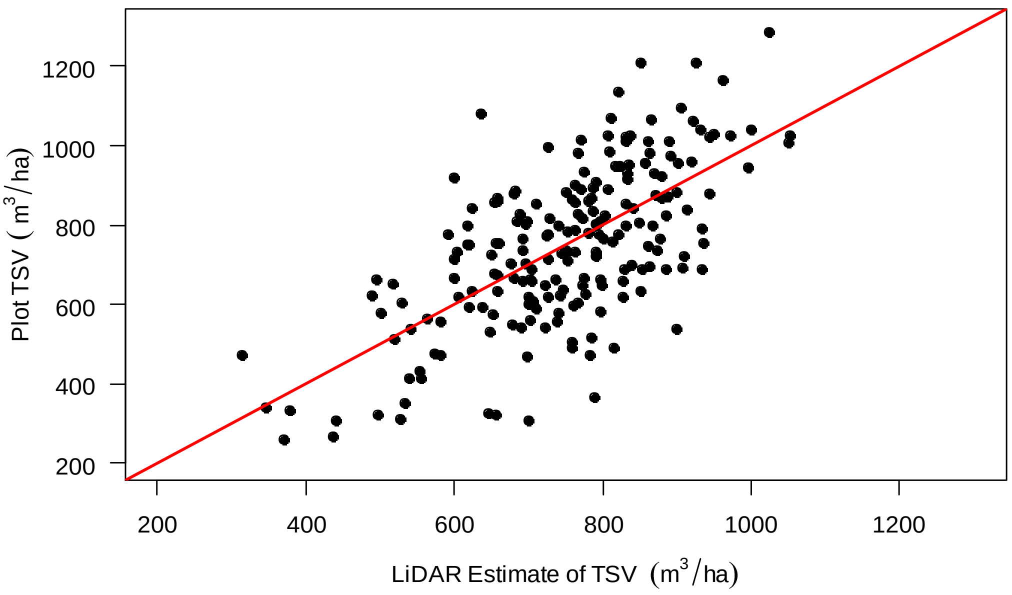

Figure 5: The relationship between the LiDAR estimate of plot TSV and the actual plot TSV as derived from the regression model which used the actual plot locations. Red line = the 1:1 relationship.

The residual analysis for this model (Equation 3) is provided in Figure 6. The model residuals ranged between -446.0 m^3/ha and 421.5 m^3/ha. Only 27% of the plot TSV residuals were within ± 50 m^3/ha and 17% of residuals were greater than ± 200 m^3/ha. The comparison between the fitted and residual values shows little bias or heteroscedasticity.

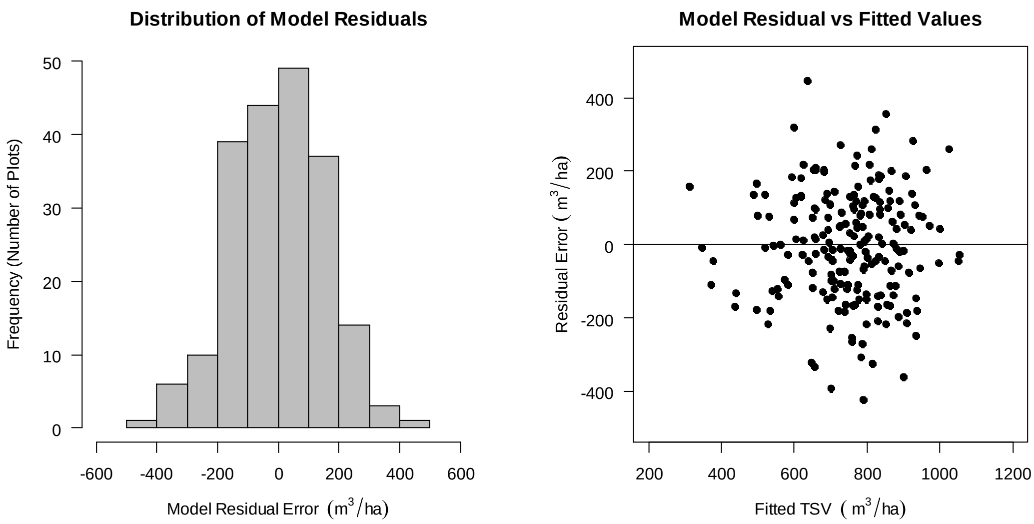

Figure 6: Analysis of the residuals for the linear model developed from actual plot location canopy metrics.

The optimal linear regression model between the measured total standing volume and the canopy metric data extracted from the intended (incorrect) locations of field plots was:

$$TSV = 503.9 + 5.10x10^{-2}(H_{20}{}^{1.95/2.95}) + 2.94x10^{-9}(PCR{>}3^{3.95/4.95}) \tag{4}$$

Notably, this model had a different intercept, slope coefficients and scaled power transformations required for the H_{20} and PCR>3 independent variables, compared to the model developed from the actual plot locations. This is because in this model (Equation 4) a lambda value of 2.95 was used for the H_{20} metric and a lambda value of 4.95 was used for the PCR>3 metric. The model coefficients of this model were all statistically significant, with p-values of less than 0.001 for the intercept and the transformed H_{20} metric, as well as a p-value of 0.006 for the transformed PCR>3 metric. Overall, the model was statistically significant, with a p-value of less than 0.001. However, the model adjusted R^2 was 19%.

The relationship between the LiDAR estimate of TSV (as derived from Equation 4) and the actual TSV for each plot is presented in Figure 7. This shows a weak relationship, where the variability in the actual plot TSV was inadequately explained by the regression model. The residual standard error was 180.0 m^3/ha. This indicates a lower accuracy in the models estimation of plot TSV, compared to the model that was developed from the actual plot locations.

The regression model developed from the intended plot locations was also less suitable for explaining the full range in the actual TSV of the plots sampled. This is because the measured total standing volumes for the sample of plots ranged from 261 m^3/ha to 1,285 m^3/ha, while the TSV estimates from Equation 4 (which was affected by the plot co-registration errors) ranged from 504 m^3/ha to 934 m^3/ha for the same sample. This is a notably tighter range and illustrates a reduced ability of the model to accurately predict the TSV of plots which had very high or low standing volumes. Comparatively, Equation 3 which was developed from the actual plot locations better explained the full range in the true plot TSV's for the sample, as its estimates ranged from 313 m^3/ha to 1,053 m^3/ha.

The residual analysis for Equation 4 is provided in Figure 8. The distribution of model residuals indicates that the error in the plot level estimate of TSV ranged between -471.5 m^3/ha and 477.4 m^3/ha. Only 20% of the plot level TSV residuals were within ±50 m^3/ha and 25% of residuals were greater than ±200 m^3/ha. The comparison between the fitted and residual values shows little bias or heteroscedasticity, but demonstrate the greater spread in the residual error. This analysis indicates that plot co-registration errors substantially reduced the strength and accuracy of regression model used to estimate the measured TSV of plots from LiDAR canopy metrics.

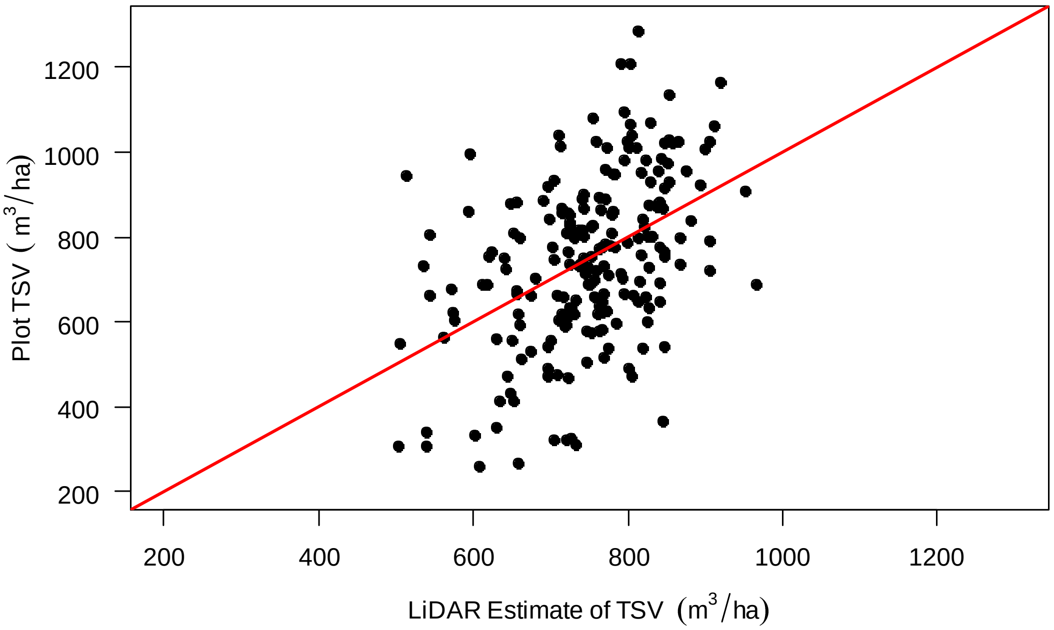

Figure 7: The relationship between the LiDAR estimate of plot TSV and the actual plot TSV as derived from the regression model which used the intended plot locations. Red line = the 1:1 relationship.

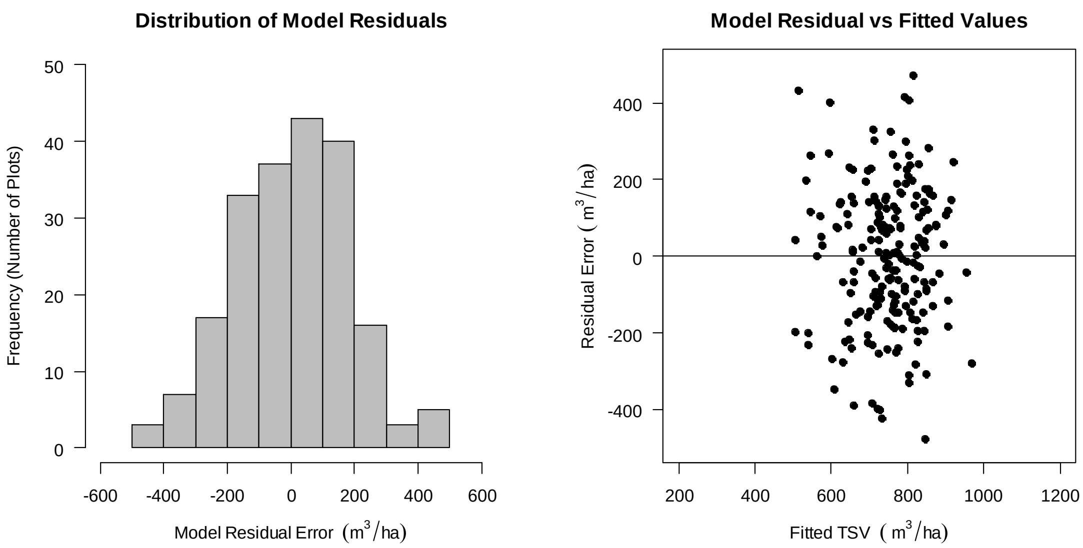

Figure 8: Residual analysis for the linear model developed from intended plot location canopy metrics.

6.3. <u>Model Strength after Stratification of the Plot Co-registration Errors:</u>

This section compares strength of regression between the transformed H_{20} and PCR>3 LiDAR canopy metrics and the actual TSV of plots, after the plots sampled had been stratified by low to high levels of plot co-registration error. This is illustrated in Figure 9, which compares the LiDAR-based estimate of TSV and the actual plot TSV, as derived from linear models that were produced for each of the five stratified groups, using either the actual ('no plot co-registration error') or intended (incorrect) locations for each plot.

This indicates that within each group, the correlation between the LiDAR estimate of TSV and the actual plot TSV was either similar or much lower in the models that were developed using the intended locations of plots, rather than the actual plot locations. This provides further evidence for the adverse effect that plot co-registration errors have on model strength and accuracy. But importantly, the correlation between the LiDAR estimate of TSV and the actual plot TSV was substantially lower in the regression models were developed using the intended plot locations that had the greatest levels of plot co-registration error, when compared to the models that were developed using the actual locations for the same sample of plots. This indicates that the increasing plot co-registration errors substantially reduced the accuracy of the regression models formed.

This analysis is supported by Figure 10, which provides a summary of the model adjusted R^2 values for these stratified regression models, with and without the plot co-registration errors. The Figure demonstrates that there was a close correspondence between the R^2 values of the models that were developed using either the intended or actual plot locations, for both the 80%-100% and the 60%-80% overlap groups. This illustrates that the low plot co-registration errors (whereby there was a high overlap between the intended and actual plot locations) had a low effect on model strength.

However, the adjusted R^2 values diverged markedly between the two types of regressions as the percentage of overlap between the actual and intended plot boundaries declined below the 60% overlap threshold (Figure 10). As the models which were developed from plots which had only a 40%-60%, 20%-40% or 0%-20% common overlap area between their actual and intended positions, had substantially lower adjusted R^2 values compared to their 'no plot co-registration error' reference points. This further demonstrates that the higher plot location errors caused greater reductions to the strength of the LiDAR-based regression models.

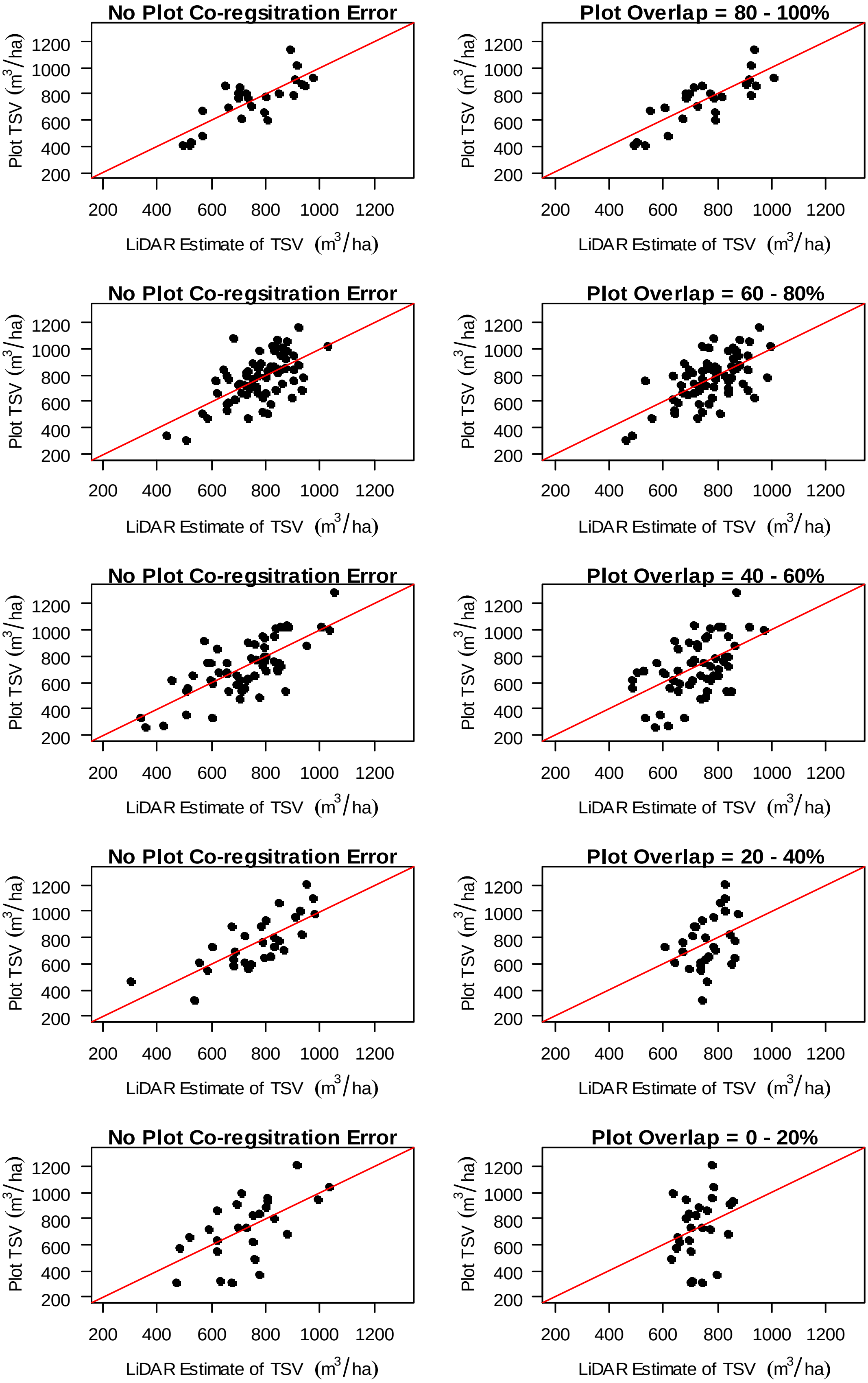

Figure 9: Comparison of model accuracy after stratification by the level of plot co-registration error (left side = the models formed from the actual plot positions, right side = the models formed from the intended plot positions). Red line = the 1:1 relationship.

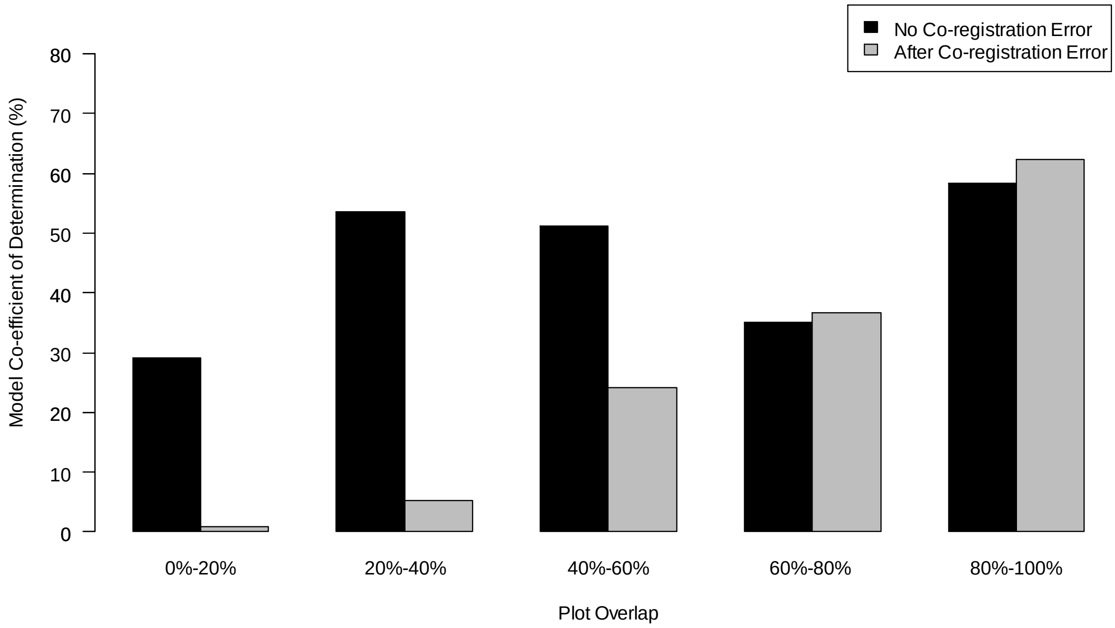

Figure 10: Variation in the model co-efficient of determination (R^2) between the stratified groups, as developed with (grey bars) and without (black bars) plot co-registration errors.

6.4. <u>The Effect of Plot Co-registration Error on Model LiDAR Canopy Metrics:</u>

This section demonstrates how the individual LiDAR canopy metrics that were used in the regression models to predict plot TSV, were affected by the plot co-registration errors. This was achieved by comparing the differences between the LiDAR metrics that were calculated in the actual and intended positions for a plot, to establish how the co-registration errors changed the LiDAR metrics that were used in the regression models to predict TSV.

Figure 11 demonstrates the absolute change that occurred to the 95th percentile elevation metric for each plot (here on referred to as H_{95}), compared to the level of plot co-registration error between the intended and actual plot locations. This metric was used to estimate the mean top height of the plots and so was indirectly included in the regression model. Figure 11 shows that as the level of co-registration error increased (represented by a decline in the overlap between the intended and actual plots) the absolute change to the H_{95} metric increased markedly. The relationship between the level of plot co-registration error and the absolute change to the H_{95} metric, as determined from the transformed variable regression analysis (Figure 20, Appendix 3), was significant (R^2 = 19%, p-value > 0.001). The outlier identifiable (co-registration error = 77% and change in metric = 4.3 m) was a result of a substantial difference in the canopy structure of the intended and actual plots. This is because the intended plot position was located within a canopy gap, while the actual plot position was offset into an area of taller closed canopy forest.

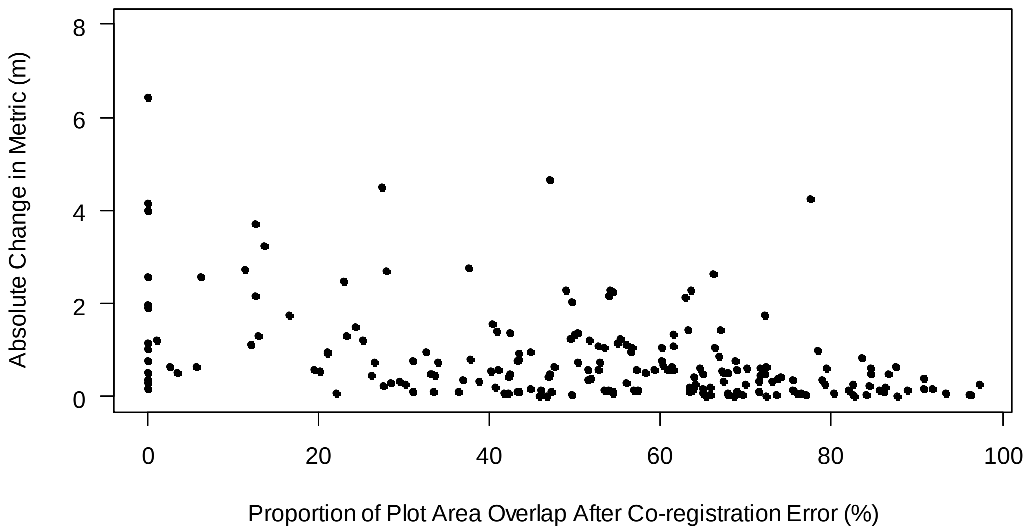

Figure 11: The relationship between co-registration error and the absolute change in the H_{95} metric.

Table 2 provides a comparison of the mean and 95% confidence intervals for the absolute and proportional differences in the H_{95} metric that occurred within each of the five stratified groups of plot co-registration error. This further indicates that as the level of co-registration error increased, the average absolute and proportional change to the H_{95} metric that was calculated for the plots also increased.

Table 2: The average absolute and percentage change (and 95% confidence intervals) in the H_{95} metric by the stratified co-registration error groups.

	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
Absolute Change in H_{95} (unit=m)	0.23 (0.09)	0.61 (0.17)	0.81 (0.21)	0.93 (0.37)	1.83 (0.58)
% Change in H_{95}	0.66% (0.26%)	1.71% (0.49%)	2.41% (0.68%)	2.66% (1.11%)	4.92% (1.62%)

The results from an Tukey's HSD analysis (Table 3) for the average absolute height changes to the H_{95} metric within each of the five stratified co-registration error groups indicates that there was a significant difference between the change in the H_{95} metric in the 0%-20% overlap group compared to all other groups and a significant difference between the change in the 20%-40% group when compared with the 80%-100% group. This indicates that higher plot co-registration errors caused significantly greater changes to the H_{95} metric, compared to lower plot co-registration errors.

Table 3: Tukey's HSD *p*-values comparing the change in the H_{95} metric between the stratified co-registration error groups. The groups with significant differences are shown in bold.

	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
80-100% Overlap		0.368	0.052	**0.033**	**0.001**
60%-80% Overlap	0.368		0.703	0.464	**0.001**
40%-60% Overlap	0.052	0.703		0.974	**0.001**
20%-40% Overlap	**0.033**	0.464	0.974		**0.002**
0%-20% Overlap	**0.001**	**0.001**	**0.001**	**0.002**	

Figure 12 demonstrates that the range in the abosolute change to the H_{20} metric (used in the regression models to predict TSV) also increased with higher plot co-regsitration errros. Importantly, the relationship between the level of plot co-registration error and the change to the H_{20} metric, as determined from the transformed variable regression analysis (Figure 21, Appendix 3) was significant (R^2 = 34%, p-value > 0.001).

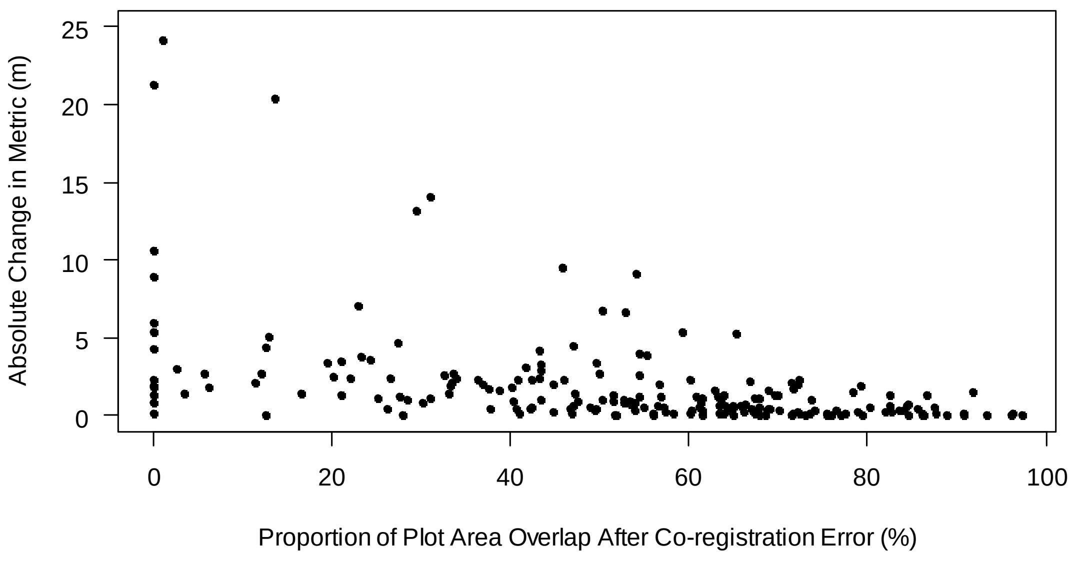

Figure 12: The relationship between co-registration error and the absolute change in the H_{20} metric.

Table 4 provides a comparison of the mean and 95% confidence intervals for the absolute and proportional differences in the H_{20} metric that occurred within each of the five stratified co-registration error groups used in the regression analysis. Notably, the changes to the H_{20} metric after the plot co-registration errors were substantially higher compared to the actual height and percentage changes in the H_{95} metric (Table 2).

Table 4: The average absolute and percentage change (and 95% confidence intervals) in the H_{20} metric by the stratified co-registration error groups.

	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
Absolute Change in H_{20} (unit=m)	0.38 (0.17)	0.78 (0.21)	1.88 (0.55)	2.96 (1.19)	5.32 (2.55)
% Change in H_{20}	1.88% (0.90%)	3.73% (1.01%)	10.32% (3.78%)	15.58% (10.09%)	23.91% (10.78%)

The Tukey's HSD analysis (Table 5) indicates that there was a significant difference in the changes to this metric between the 0%-20% overlap group with all other groups and a significant difference between the 20%-40% group with the 80%-100% group and the 60%-80% group. This further shows how the higher plot co-regsitration errors caused greater changes to the H_{20} metric.

Table 5: Tukey's HSD *p*-values comparing the change in the H_{20} metric between the stratified co-registration error groups. The groups with significant differences are shown in bold.

	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
80-100% Overlap		0.979	0.204	**0.013**	**0.001**
60%-80% Overlap	0.979		0.221	**0.009**	**0.001**
40%-60% Overlap	0.204	0.221		0.486	**0.001**
20%-40% Overlap	**0.013**	**0.009**	0.486		**0.026**
0%-20% Overlap	**0.001**	**0.001**	**0.001**	**0.026**	

Last, the change in the PCR>3 Metric after the plot co-registration errors, is presented in Figure 13. The change is recorded on a percent basis because this metric defines the proportion of point returns above 3.00 metres in a LIDAR point cloud. This also shows an increasing trend in the absolute change to this metric with higher plot co-registration errors. The relationship between the level of plot co-registration error and the change to the PCR>3 metric, as determined from the transformed variable regression analysis (Figure 22, Appendix 3), was significant (R^2 = 22%, *p*-value > 0.001). This trend is supported by Table 6, which indicates that the change in the PCR>3 metric increased with greater levels of plot co-registration errors, on both an absolute and percentage of original value basis.

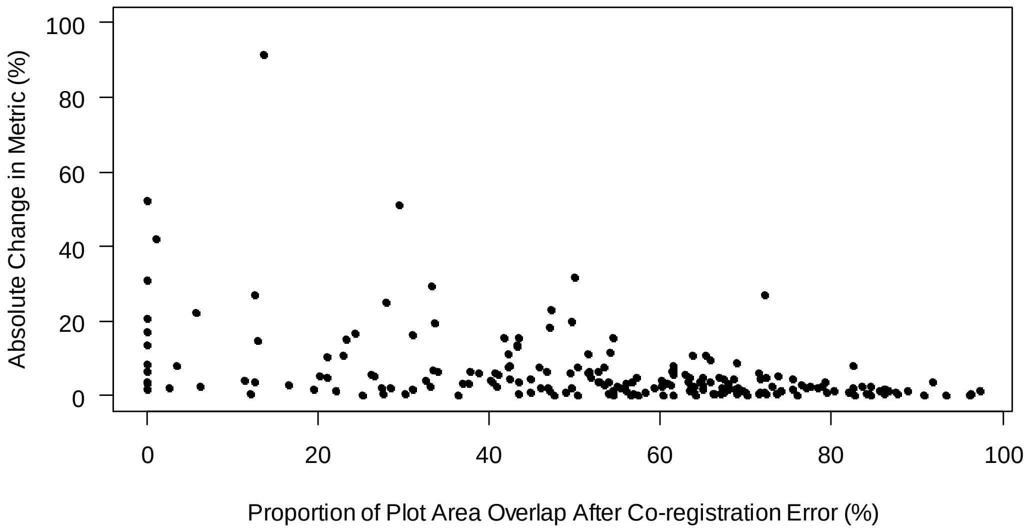

Figure 13: The relationship between co-registration error and the absolute change in the PCR>3 metric.

Table 6: The average absolute and percentage change (and 95% confidence intervals) in the PCR>3 metric by the stratified co-registration error groups.

	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
Absolute Change in PCR>3 (unit = %)	2.51 (1.06)	6.05 (1.47)	10.78 (2.66)	14.51 (5.07)	26.93 (13.90)
% Change in PCR>3	1.42% (0.66%)	3.49% (0.95%)	6.22% (1.65%)	9.13% (4.04%)	15.29% (7.87%)

The Tukey's HSD analysis (Table 7) also indicates that there was a significant difference in the change to the PCR>3 metric between the 0%-20% overlap group with all other stratified groups, and a significant difference between the 20%-40% group with the 80%-100% group. This indicates that the plot co-registration errors also affected the point density metrics, such as PCR>3 by a similar extent compared to the LiDAR canopy height metrics, such as H_{95} and H_{20}.

Table 7: Tukey's HSD *p*-values comparing the change in the PCR>3 metric between the stratified co-registration error groups. The groups with significant differences are shown in bold.

	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
80-100% Overlap		0.878	0.200	**0.023**	**0.001**
60%-80% Overlap	0.878		0.477	0.054	**0.001**
40%-60% Overlap	0.200	0.477		0.639	**0.001**
20%-40% Overlap	**0.023**	0.054	0.639		0.106
0%-20% Overlap	**0.001**	**0.001**	**0.001**	0.106	

This analysis has demonstrated that increasing plot co-registration errors (where by intended and actual location of plots become further separated in the LiDAR point cloud) significantly changed the LiDAR canopy metrics that were estimated for each plot and used in the regression model with TSV.

6.5. The Effect of Plot Co-registration Error on the Percentile Elevation Metrics

Figure 14 compares the absolute changes (in metres) that occurred to the differing LiDAR percentile elevation metrics, by the level of co-registration error between the intended and actual locations for each plot. These canopy metrics quantify the height of point returns in the point cloud at threshold levels. This indicates that for all of the LiDAR canopy percentile elevation metrics, increasing plot co-registration errors (as indicated by a reduction in the plot overlap) resulted in greater absolute changes to their values.

However, there was a difference in the magnitude of response that each canopy height metric had to the plot co-registration errors. This is because the range in the absolute changes to these metrics was higher for the metrics which were calculated at the lower percentile levels of the point cloud (such as the 10th and 20th percentile elevations), compared to the metrics calculated from the higher percentile levels (such as the 80th and 95th percentile elevations). This is supported by Table 8, which provides a numerical comparison for the mean absolute changes in height for the differing LiDAR percentile elevation metrics, by the five stratified groups of plot co-registration error.

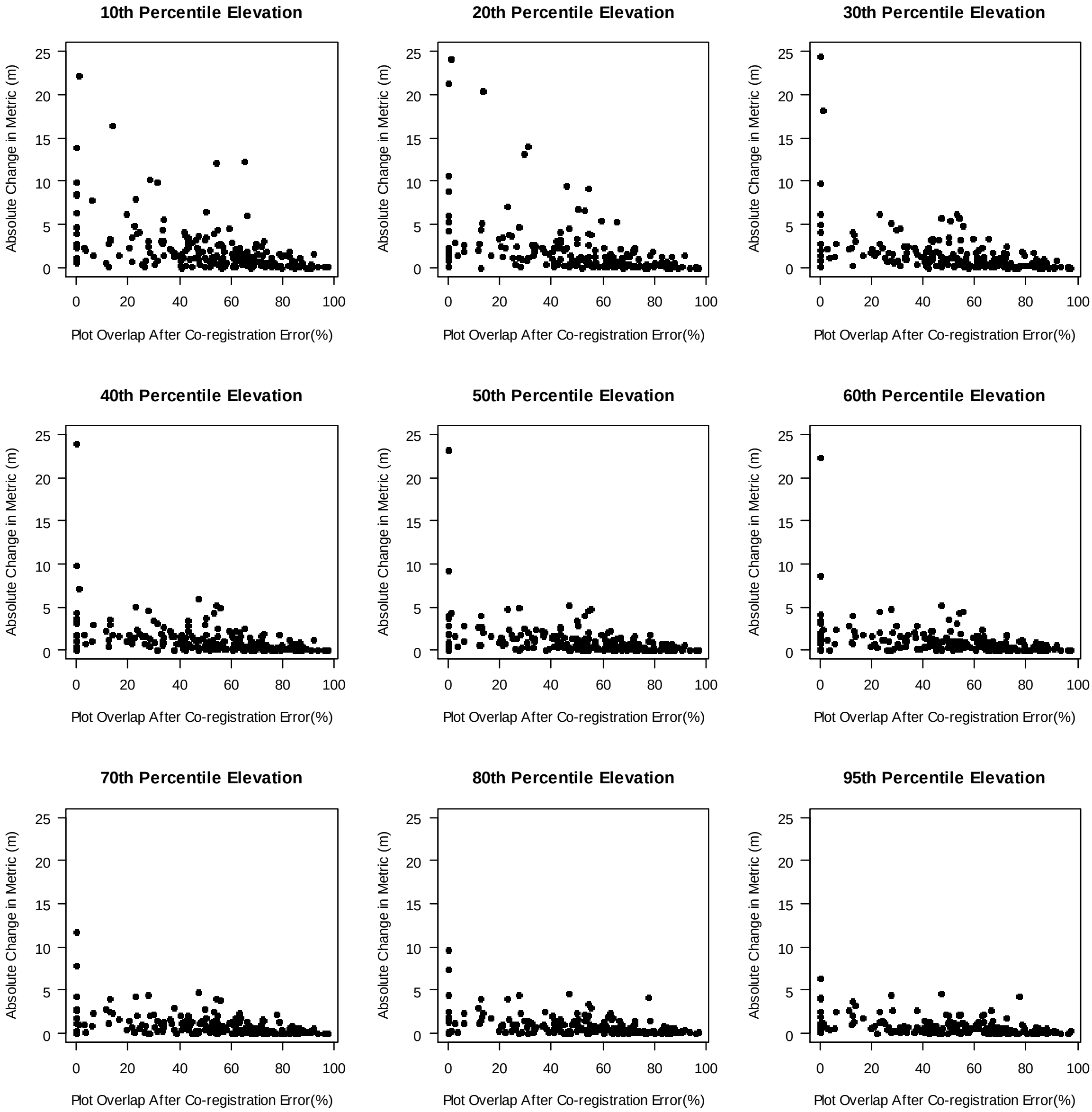

Figure 14: The relationship between co-registration error and the absolute changes to the different LiDAR percentile elevation metrics.

Table 8: The absolute average metre change in the different percentile of elevation LiDAR metrics by the stratified co-registration error groups.

Stratified Group	Percentile of Elevation Above the Ground Level Metric									
	10^{th}	20^{th}	30^{th}	40^{th}	50^{th}	60^{th}	70^{th}	80^{th}	90^{th}	95^{th}
80%-100% Overlap	0.50	0.38	0.37	0.34	0.30	0.30	0.28	0.28	0.25	0.23
60%-80% Overlap	1.21	0.78	0.66	0.66	0.64	0.58	0.58	0.59	0.59	0.61
40%-60% Overlap	1.93	1.88	1.62	1.32	1.22	1.16	1.08	0.96	0.89	0.81
20%-40% Overlap	3.05	2.96	1.99	1.72	1.52	1.41	1.22	1.09	0.95	0.93
0%-20% Overlap	5.25	5.32	4.18	3.36	3.11	2.81	2.24	2.07	1.87	1.83

Expressing these changes as a percentage of the original elevation above the ground level accentuates these patterns. Figure 15 demonstrates the relationship between the percentile of elevation above the ground level for each height metric and the mean proportional change to each metric, within each of the five stratified groups of plot co-registration errors. This indicates that for each height metric, the mean change in their values was higher in the stratified groups that had greater levels of plot co-registration error. But within each stratified group, there was an exponential decline in the proportional change in the metrics as their percentile height above the ground level increased. Such that, the lower metrics were much more sensitive to the plot co-registration errors, compared to the higher level metrics. As an example, the mean change in the 10th height percentile metric after plot co-registration errors that caused either an 80-100% or 0-20% plot overlap was 3% and 36%, respectively. In comparison, the mean change in the 95th height percentile metric for these locational errors was 1% and 5%, respectively.

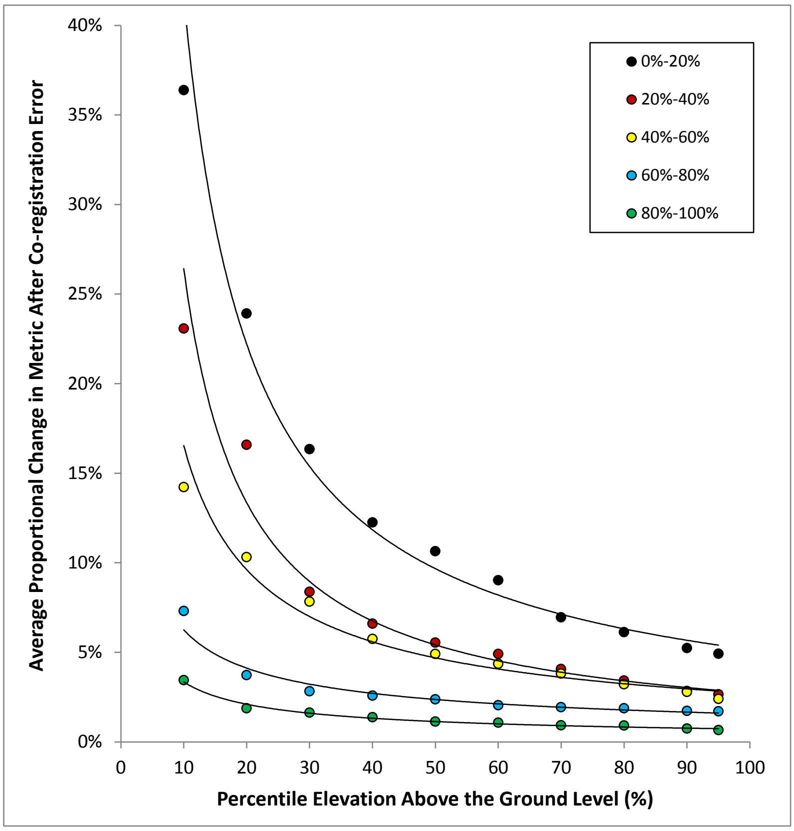

Figure 15: The relationship between the percentile of elevation above the ground level for each metric and the mean % change of each height metrics, within the stratified groups. Green circles = 80%-100% overlap group, blue circles = 60%-80% overlap group, yellow circles = 40%-60% overlap group, red circles = 20%-40% overlap group and black circles = 0%-20% overlap group.

7. DISCUSSION

7.1. Explanation of the Results and Implications:

This study has demonstrated that the management decision to not record the actual position of inventory field plots and to instead use the intended position to define the location of these field plots caused a significant plot co-registration error. Whereby, it was evident that this plot co-registration error was an important consideration to the LiDAR-based forest inventory in the plantation of mature *Pinus radiata.* This was demonstrated through the three main findings, which were that:

1. The strength and accuracy of the regression models formed between LiDAR canopy metrics and the measured total standing volume of plots reduced substantially after increasing plot co-registration errors.
2. The LiDAR canopy metrics that were used as independent variables in the regression models changed significantly after increasing plot co-registration errors between the actual and intended positions for plots.
3. The LiDAR canopy metrics that were calculated at the lower elevations of the LiDAR point cloud were more sensitive to change after plot co-registration errors, compared to the metrics which were calculated from the higher elevations.

The consideration of these findings demonstrates the importance of accurately locating plot centres in the LiDAR based pre-harvest inventory. This was because the correct definition of plot locations (through measurement with the survey grade GPS) resulted in the use of LiDAR data that was spatially synchronous to the inventory plots that were measured in the trial forest. Whereby, the application of this LiDAR data to the regression model with TSV produced a higher model strength and accuracy.

In comparison, the plot co-registration errors (which were caused by the inaccurate procedure to define plot locations) reduced the strength of regression between LiDAR canopy metrics and TSV. This occurred because the LiDAR canopy metrics had a high sensitivity to change after the plot co-registration errors. This caused the LiDAR metrics that were calculated from the intended position for a plot to be different from the LiDAR metrics that were calculated from the plots actual location. This is significant, as it resulted in an invalid regression between the LiDAR data and TSV after the plot co-registration errors; because the LiDAR metrics which were extracted from the incorrect locations of each plot were unrepresentative of the trees that were measured in the forest inventory. This erroneous relationship reduced the strength of the regression models.

Importantly, the adverse effect of the plot co-registration errors on the strength of regression was greater as the level of locational error increased. This would be a result of the increasing absolute change that occurred to the LiDAR canopy metrics (which were used in the regression models) after the higher plot location errors. Notably, the reduction in model strength was most substantial after there was less than a 60% plot area overlap between the intended and actual positions for a plot. This overlap threshold equated to a minimum linear co-registration error distance of 7.2 m to 8.7 m for the plots in the sample which ranged from 400 m^2 to 600 m^2 in size (Table 9, Appendix 4).

The differentiated response of LiDAR percentile height metrics (which were calculated at varying elevations of the forest canopy) to the plot co-registration errors also has a critical implication to the LiDAR-based forest inventory. This is because lower level LiDAR canopy metrics, such as the 20^{th} and 30^{th} percentile of elevation metrics, are commonly used in regression models to predict TSV in operational LiDAR-based forest inventories (Marshall et al., 2013; Watt & Watt, 2013). Importantly, this study has demonstrated that these metrics will be highly sensitive to change after the plot location errors. Therefore, plot co-registration errors will be a particularly important consideration to the strength of LiDAR-based regressions that are used to estimate forest biomass or yield. In comparison, the strength of the regression models that are used to estimate the mean top height of plots from metrics such as H_{95} would be less affected by plot co-registration errors, given the lower sensitivity of the higher level metrics to plot location errors.

These results indicate that implementing procedures to accurately define field plot locations will be an important requirement for the LiDAR-based pre-harvest inventory in New Zealand forests of *Pinus radiata*. This is because this will eliminate the adverse effect of plot co-registration errors on the strength of regression formed between LiDAR canopy metrics and measured stand inventory variables. This could be achieved by using survey grade GPS units to measure plot locations, as implemented in this study. Furthermore, the use of larger plot sizes in the LiDAR-based forest inventory could also be beneficial, as this will reduce the relative effect of linear plot locational errors.

7.2. Critical Evaluation of the Findings:

This study has provided a verification of the effect that plot co-registration errors have on the strength of regressions between LiDAR canopy metrics and total standing volume, which aligns with the findings in the research by Gobakken and Naesett (2009) and Frazer et al (2011). This is significant, as it demonstrates that plot co-registration errors can also be a limitation to the LiDAR-based forest inventory which is conducted in the homogenous forest type of *Pinus radiata*, as similar to the heterogeneous forest types. This is although Marshall (2012) had demonstrated a relatively low sensitivity of the *average* model strength to plot co-registration errors in *Pinus radiata* forests, while Gobakken and Naesett (2009) had suggested that the LiDAR-based forest inventories which are implemented in even forest structures would be less affected by plot positional errors.

Arguably, the results observed in this study would have occurred because, although the forest was categorised as homogenous at the stand level, there would have been a high variability in the structure of the *Pinus radiata* canopy throughout the stands sampled. This occurred because of the high incidence of wind throw and mortality across the trial forest, as well as because of the differences in the structure of the forest canopy which occurred along differing positions of the ground terrain and at the canopy edges.

This variability in the structure of the forest canopy is significant as it would have influenced the effect of the plot co-registration errors on the regression strength between LiDAR canopy metrics and TSV. This is because research has identified that the three-dimensional form of the LiDAR point cloud is highly sensitive to changes in the structure of the forest canopy (Kane et al., 2010). Hence, the variability in the canopy structure of the trial forest would have caused the changes to the LiDAR metrics after the plot co-registration errors. As the canopy structure would have differed between the intended and actual positions for a plot. This in turn resulted in the development of the in-valid regressions, after the plot co-registration errors, which reduced the model strength achieved. Furthermore, the adverse effect of the plot co-registration errors occurred to a greater extent after the highest plot locational errors, which aligns to the expectation that there would have been greater changes to structure of forest canopies after the increasing spatial errors.

Therefore, because of the variability in the structure of the *Pinus radiata* forest canopy, it will be essential to minimise plot location errors in the LiDAR-based forest inventory. As this will eliminate the potential for incorrect and unrepresentative LiDAR canopy metrics to be applied to the regression models that are used to predict stand variables.

However, it is also important to recognise that there was also a degree of homogeneity in the canopy structure of the trial *Pinus radiata* forest, which acted as a natural buffer against large changes in plot-level metrics after the plot co-registration errors. This is because in some cases, the highest plot co-registration errors only resulted in small changes to the LiDAR metrics that were used in the regression models. This would be due to the similarity in the structure of plots across these distances. Therefore, the homogeneity that was present between some areas of the forest canopy could have prevented a greater reduction to the strength of the regression models that were produced after the plot co-registration errors. Notably, this natural buffer from the *Pinus radiata* forest canopy could explain why Marshall (2012) did not record a substantial change in the average regression strength between TSV and the 30th percentile elevation LiDAR canopy metric after increasing plot co-registration errors.

Finally, another key feature of the point cloud is that the density of point returns decreases with a reduced vertical elevation above the ground level. This would be a result of the interception of the aerially emitted LiDAR pulses by the upper canopy of the forest, which would cause an under representation of point returns in the lower region of the plots. This feature is evident in Figure 16, which shows four LiDAR point clouds that were extracted from the boundaries of four plots. This is important, because it indicates that the statistical metrics which were calculated from the lower levels of the point cloud (such as H_{20}) would be derived from a much lower sample size (the number of point returns per plot), compared to the higher metrics (such as H_{95}). Because of this, the lower level metrics would have been more prone to greater changes after the plot co-registration errors. As an alteration to the vertical distribution of point returns in a plot, which would occur after a plot co-registration error, would be more likely to have a greater effect on the metrics which describe the lower percentiles of point returns, compared to the higher percentile metrics. This could explain why there was a differentiated response in the LiDAR percentile height metrics to the plot co-registration errors. However, there is currently limited evidence to support this hypothesis, because this effect has not yet been comprehensively reviewed in any other study.

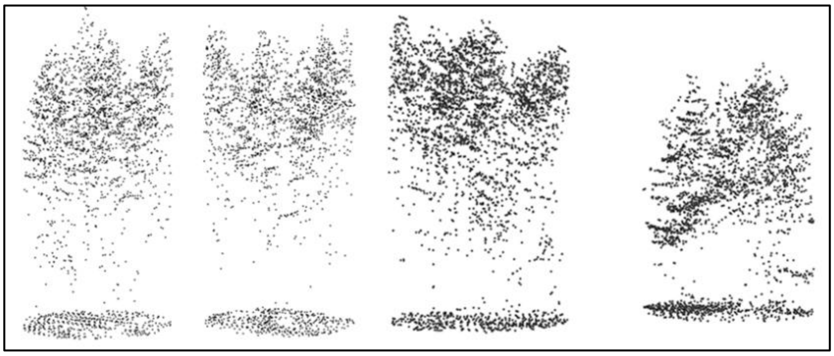

Figure 16: A comparison of LiDAR point clouds extracted from Waipapa forest.

7.3. Future Research:

It is recommended that future research further examine how the variability in the structure of forest canopies influences the derivation and changes to LiDAR canopy metrics after plot co-registration errors. Specifically, this research could be applied to develop predictive models to estimate the expected change to LiDAR canopy metrics after plot co-registration errors. Factors such as; the size of plots used in the forest inventory and the LiDAR canopy metrics which describe the density and variability of point returns could be used as independent variables in these models. These models could be used to determine when plot location errors would substantially reduce the strength of LiDAR-based regressions. Additionally, further research may be required to provide a validation for the differentiated response of LiDAR percentile elevation metrics to plot co-registration errors in other forest settings.

7.4. Study Limitations:

It is also important to consider the limitations of this study. The first of which was that the experimental design of this study was uncontrolled. This is because each of the 204 plots that were sampled was affected by a different level of plot co-registration error and hence the effect of the plot co-registration error was not equal for all of the plots. This could have caused interacting factors to have interfered with the effect of the plot co-registration errors on the strength of regression recorded in this study. This would reduce the ability to interpret the specific effect of the plot co-registration errors to the accuracy gained from the LiDAR-based forest inventory.

This limitation was a direct result of the premise of this study, which was to assess the effect of plot co-registration error in an actual forest inventory operation. However, in order to have provided a more effective analysis, a controlled experimental design should have been used. An analysis procedure similar to the study by Marshall (2012) would have been advantageous.

Another limitation in the study was that the measure of plot total standing volume, which was used as the dependant variable in the regression model, could have been incorrect. This could have occurred because of the temporal co-registration error that existed between the timing of the forest inventory (which occurred between 2009 and 2013) and the date of LiDAR acquisition for the forest (August 2013). This resulted in the requirement to simulate the growth of the plot level stand inventory variables to the date of LiDAR measurement. If the growth models used were incorrect, or if factors such as wind-throw had affected plots within this time period, then the stand variables that were assumed correct for each plot would not adequately represent the actual conditions in each plot at the date that the LiDAR point cloud was produced for the forest.

This would influence the ability to develop strong relationships between the LiDAR canopy metrics and the total standing volume of plots, because it would result in the application of incorrect estimates of TSV to the LiDAR canopy data. This could cause an invalid relationship in the regressions of this study because the LiDAR canopy data would inadequately represent the stand variables, regardless of plot co-registration error. This could explain why the regression strength between the LiDAR canopy metrics and TSV was relatively low for the linear model that was developed using the actual plot locations.

Another limitation was that there could also have also been a plot co-registration error in the actual plot locations, which were used to define the ‘no co-registration error’ reference points of this study. This is because survey grade GPS unit used to measure the plot centres could have been inaccurate, as these units do not always have sub-meter positional accuracy under forest canopies. Weih et al., 2009 indicated that the post processed locations derived from the Trimble Geo XH, a survey grade GPS, had an accuracy of 3.92 metres accuracy under canopy cover. Additionally, this spatial error could have been increased by the imperfect ability to accurately define the locations of plot centres in the field. This is because the procedures that were applied to identify the centre of plots could also have been inaccurate; as the centre peg was not always visible within all plots, the plot centre tree did not always adequately indicate the geometric centre of a plot and the process of manually determining the geometric centre of the plot through visually surveying the measured trees was inaccurate.

This is a major limitation to this study, as it would reduce the ability to demonstrate how model strength is impacted by plot co-registration errors. This is because the model that was assumed to have ‘no plot co-registration error’ could have been affected by a slight co-registration error. This would have reduced the strength of the optimal model and would reduce the ability to demonstrate how co-registration error affected model strength, when compared to the ‘no plot co-registration error’ reference points.

Finally, the results of this study are only specific to the conditions and analysis procedure that was undertaken for the trial forest. The study was only conducted in the forest structure of a mature (aged 24-27), pruned, *Pinus radiata* stand located in the Northern Hawkes Bay on low elevations. Alternative forest structures were not evaluated, so it could be possible that the results observed are not representative of other age classes, silvicultural regimes or regions. The study has also only assessed the effect of eliminating plot locational errors on the linear regression analysis procedure. The effect was not tested in context of the kNN analysis method, which is another common tool for the LiDAR-based forest inventory (Dash et al., 2013). These factors could limit the applicability of the results of this study to other forestry settings and analysis procedures. Therefore, this research should only be considered as a case study for the effect of plot co-registration errors on the LiDAR-based forest inventory.

8. CONCLUSIONS

This study has provided evidence to demonstrate that plot co-registration errors have a significant effect on the utility of LiDAR for the forest inventory. This was because a higher strength of regression was achieved between LiDAR canopy metrics and the measured TSV of plots, when the LiDAR data that was used in the regression models were spatially synchronous to the field plots that were measured in the pre-harvest inventory. As in comparison, the plot co-registration errors resulted in incorrect and unrepresentative LiDAR metrics to be derived for each plot, which caused in-valid regression models when this erroneous data was applied to predict the field measured TSV. This reduced the strength of the regression models formed. The effect of this was most notable after the plot location errors resulted in less than 60% overlap between the boundaries of the correct and incorrect locations for a plot. Furthermore, it is evident that plot locational errors will be a particularly important consideration for the LiDAR-based forest inventories which target accurate estimates of forest biomass or yield. This is because the metrics that are commonly used in these models will have a high sensitivity to change after plot co-registration errors.

Therefore, when conducting LiDAR based pre-harvest inventories in mature *Pinus radiata* plantations, the location of field plots should be accurately measured. This will avoid the adverse impact of plot co-registration errors on the strength of regression between LiDAR canopy metrics and measured stand inventory variables. This finding hence provides verification for the effect of plot co-registration errors on an actual LiDAR-based forest inventory in a *Pinus radiata* forest. This indicates that plot co-registration errors are an important consideration to the accuracy that is gained from the regression models which predict TSV using LiDAR canopy metrics in the New Zealand forestry setting.

ACKNOWLEDGEMENTS

I gratefully acknowledge the support of Dr Justin Morgenroth (New Zealand School of Forestry, University of Canterbury) and Dr Michael Watt (Scion) as supervisors of this study. Their guidance and advice has assisted me considerably during the study.

I also thankfully recognise and acknowledge Pan Pac Forest Products Ltd for their support of this study. This is because the staff of Pan Pac's Forestry and Logistics Department saw value in the project and allowed me to undertake the research. Pan Pac also provided the LiDAR and pre-harvest inventory data, funded the costs of field testing and provided guidance during the study. In particular, Brett Gilmore, of Pan Pac Forest Products Ltd, has been a fantastic mentor to the study and has also guided my personal and professional development in the forestry industry over the past three years.

REFERENCES

Adams, T., Brack, C., Farrier, T., Pontl, D., & Brownlie, R. (2011). So you want to use LiDAR? - A guide on how to use LiDAR in forestry. New Zealand Journal of Forestry 55(4), 19-23.

Dash, J., Marshall, H., & Rawley, B. (2013). Nearest Neighbour Imputation of Stand Attributes using LiDAR . Rotorua, New Zealand: Future Forests Research Ltd.

Eskelson, B. N., Temegsen, H., Lemay, V., Barrett, T. M., Crookston, N. L., & Hudak, A. T. (2009). The roles of nearest neighbour methods in inputting missing data in forest inventory and monitoring datasets. Scandinavian Journal of Forest Research, 235-246.

Frazer, G. W., Magnussen, S., Wulder, M. A., & Niemann, K. O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. Remote Sensing of Environment 115, 636-649.

Gobakken, T., & Naesett, E. (2009). Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. Canadian Journal for Forest Research 39, 1036-1052.

Johnson, C. E., & Barton, C. C. (2004). Where in the world are my field plots? Using GPS effectively in environmental field studies. Frontiers in Ecology and the Environment 2(9), 475-482.

Kane, V. R., McGaughey, R. J., Bakker, J. D., Gersonde, R. F., Lutz, J. A., & Franklin, J. F. (2010). Comparisons between field - and LiDAR-based measures of stand structural complexity. Canadian Journal of Forest Research (40), 761-773.

Kimberley, M. O., and P. N. Beets. 2007. “National Volume Function for Estimating Total Stem Volume of Pinus Radiata Stands in New Zealand.” New Zealand Journal of Forestry Science 37: 355–371.

Krogstad, F., & Schiess, P. (2004). The allure and pitfalls of using LiDAR topography in harvest and road designs. Forest Operations under Mountainous Conditions (pp. 1-10). Vancouver, B.C: IUFRO.

Marshall, H. (2012). What Accuracy / Precision is Required When Fixing Location of LiDAR Ground Controls. Retrieved from Interpine: http://www.interpine.co.nz/news/Lists/Posts/Post.aspx?ID=104

Marshall, H., Dash, J., Rawley, B., & Adams, T. (2012). Using LiDAR based Regression Estimation in New Zealand Forest Inventory. Rotorua, New Zealand: Future Forests Research Ltd.

Popescu, S. C. (2007). Estimating biomass of individual pine trees using airbourne lidar. Biomass and Bioenergy 31, 646-655.

US Forest Service. (2013). FUSION. Pacific Northwest Research Station, http://www.fs.fed.us/eng/rsac/fusion/.

Watt, M. S., Adams, T., Marshall, H., Pont, D., Lee, J., Crawley, D., & Watt, P. (2013). Modelling variation in Pinus radiata stem volume and outerwood stress-wave velocity from LiDAR metrics. New Zealand Journal of Forestry Science 43(1), 1-7.

Watt, P., & Watt, M. S. (2013). Development of a national model of Pinus radiata stand volume from LiDAR metrics for New Zealand. International Journal of Remote Sensing , 1-13.

Weih, R. C., Gilbert, M., Cross, J., & Freeman, D. (2009). Accuracy assessment of recreational and mapping grade GPS receivers. Journal of the Arkansas Academy of Science 63, 163-168.

White, J. C., Wulder, M. A., Varhola, A., Vastaranta, M., Coops, N. C., Cook, B. D., Woods, M. (2013). A best practice guideline for generating forest inventory attributes from airbourne laser scanning data using an area based approach. Victoria, British Columbia: Canadian Wood Fibre Service, Canadian Forest Service.

Wulder, M. A., Bater, C. W., Coops, N. C., Hilker, T., & White, J. C. (2008). The role of LiDAR in sustainable forest management. The Forestry Chronicle 84(6), 807-826.

APPENDICES

Appendix 1: The Locations of the Plots Sampled in the Study:

The locations of the plots which were randomly sampled are demonstrated in Figure 17 below. Importantly, plots were not re-measured in the south western corner of the forest due to time restrictions. To account for the loss, an additional 13 plots were selected from outside of the 160 metre radius around other grid intersections.

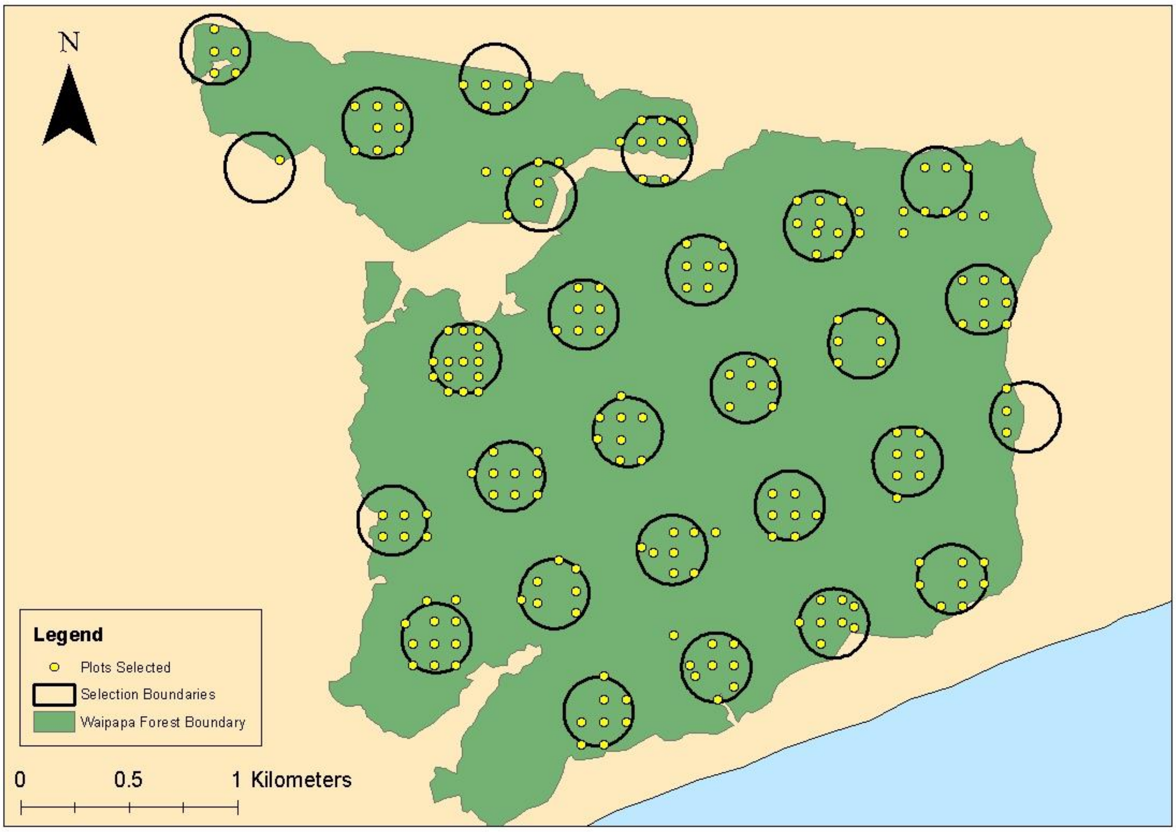

Figure 17: Plots selected for re-measurement in Waipapa Forest

Appendix 2: The Justification for the use of a LiDAR Derived MTH:

The conclusion that the field measured MTH estimates would be inadequate for use in this study was established through a comparison of the LiDAR based MTH estimate for each of the plots, to the mean top height (MTH) determined from the pre-harvest inventory. Whereby, the LiDAR based MTH estimate is derived from a function of 95Th percentile elevation canopy metric (Equation 1, Page 14). Overall the relationship was very poor (Figure 18) and indicated that the manual MTH estimates for each plot did not align with the estimates calculated from the LiDAR analysis.

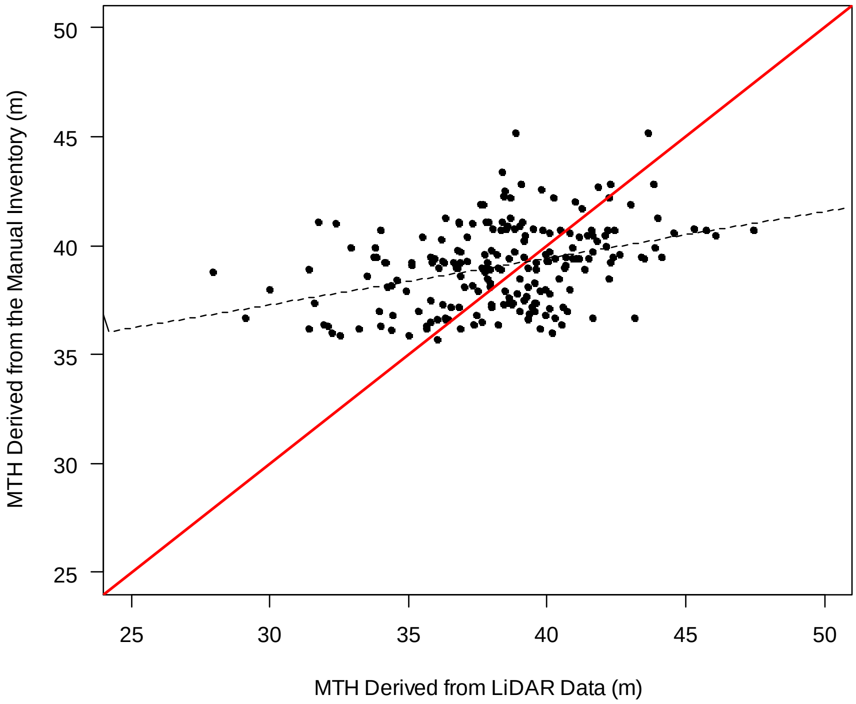

Figure 18: The relationship between the LiDAR based and manual based estimates of plot MTH. With 1:1 line (red) and the line of best fit regression (dotted).

It is postulated that the poor relationship indicates an inaccuracy in the manually based procedure used to estimate the MTH of each plot. This is because the manual MTH estimates could have been limited by the low number of trees that were measured for height within each plot and because the manual procedure of measuring the total height of trees with a vertex can be inaccurate. Furthermore, once the manually measured plot height tree data was transferred to Atlas Cruiser, the plot level MTH estimates were determined from the stand level relationship between the height and diameter of all trees that were measured for height within each stand. Whereby, each plot was allocated a MTH on the basis of the mean diameter within each plot. This function resulted in estimates of MTH that were not specific to each plot in the sample. The averaging effect from this function can be observed in Figure 18, where the manual based estimates of MTH were within a tight range (36 m to 45 m). In comparison, the LiDAR estimate of MTH within each point cloud had a wider range, as its estimates were from 28 m to 45 m.

Finally there is also evidence that as the proportion of trees measured for height in a plot increased, the absolute difference between the manual based estimate and the LiDAR based estimate of MTH reduced (Figure 19). This indicates that by measuring a greater number of trees for plot for height, the accuracy in the manual estimate of MTH increased, relative to the LiDAR based estimate of MTH. This provides further evidence to suggest that the LiDAR based estimate of MTH was more suitable for use in this study, compared to manual estimate.

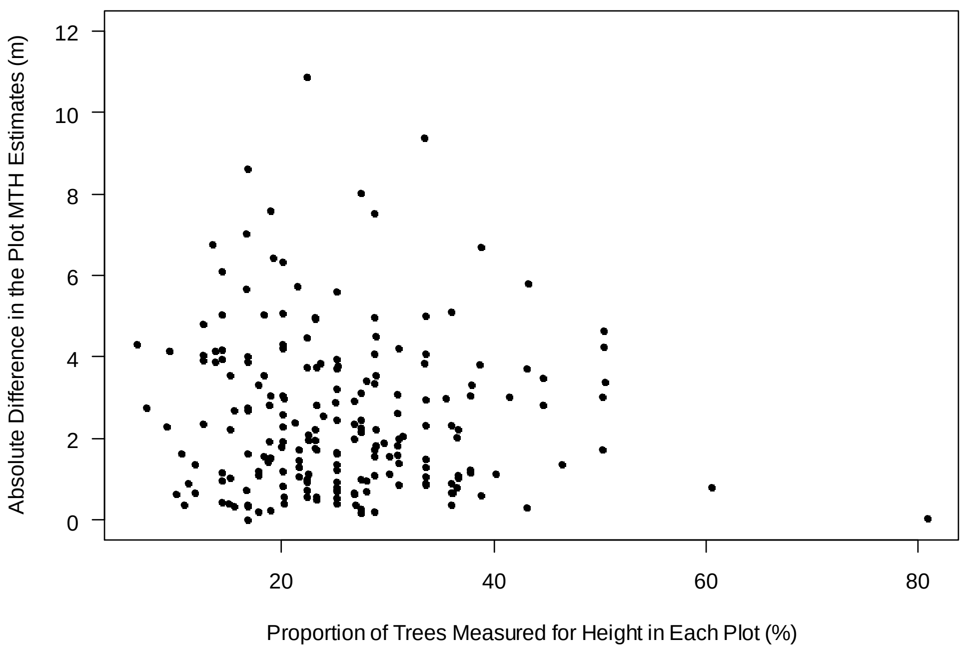

Figure 19: The difference between the LiDAR based and manual based estimates of plot MTH, by the proportion of plot trees measured for height per plot.

Appendix 3: The Transformed Relationships between the Plot Overlap Co-registration Errors and the LiDAR Canopy Metrics

The H_{95} LiDAR Canopy Metric

The relationship between the transformed H_{95} LiDAR canopy metric and the transformed plot overlap is presented in Figure 20. The regression between these two variables had the model form:

$$H_{95}^{(-0.30/1.30)} = 0.175 - (0.0034 * Overlap^{0.45/1.45})$$

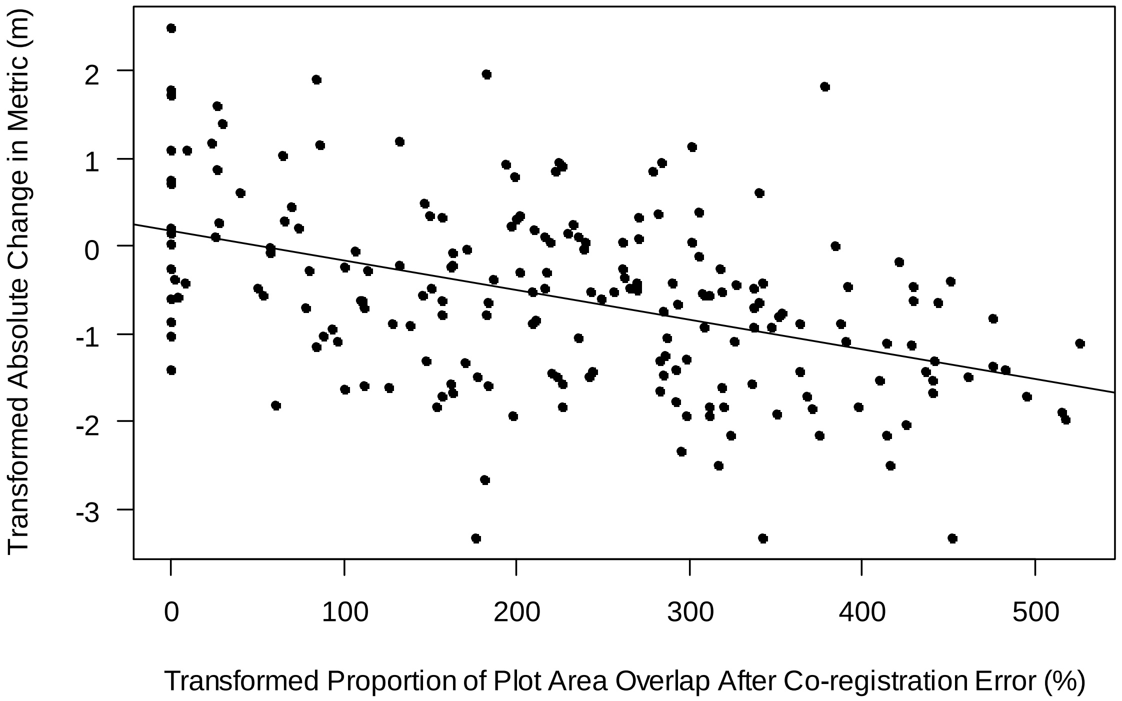

Figure 20: The relationship between the transformed H_{95} LiDAR canopy metric and the transformed plot overlap

The H_{20} LiDAR Canopy Metric

The relationship between the transformed H20 LiDAR canopy metric and the transformed plot overlap is presented in Figure 21. The regression between these two variables had the model form:

$$H_{20}^{(-0.20/1.20)} = 0.175 - (0.0034 * Overlap^{0.45/1.45})$$

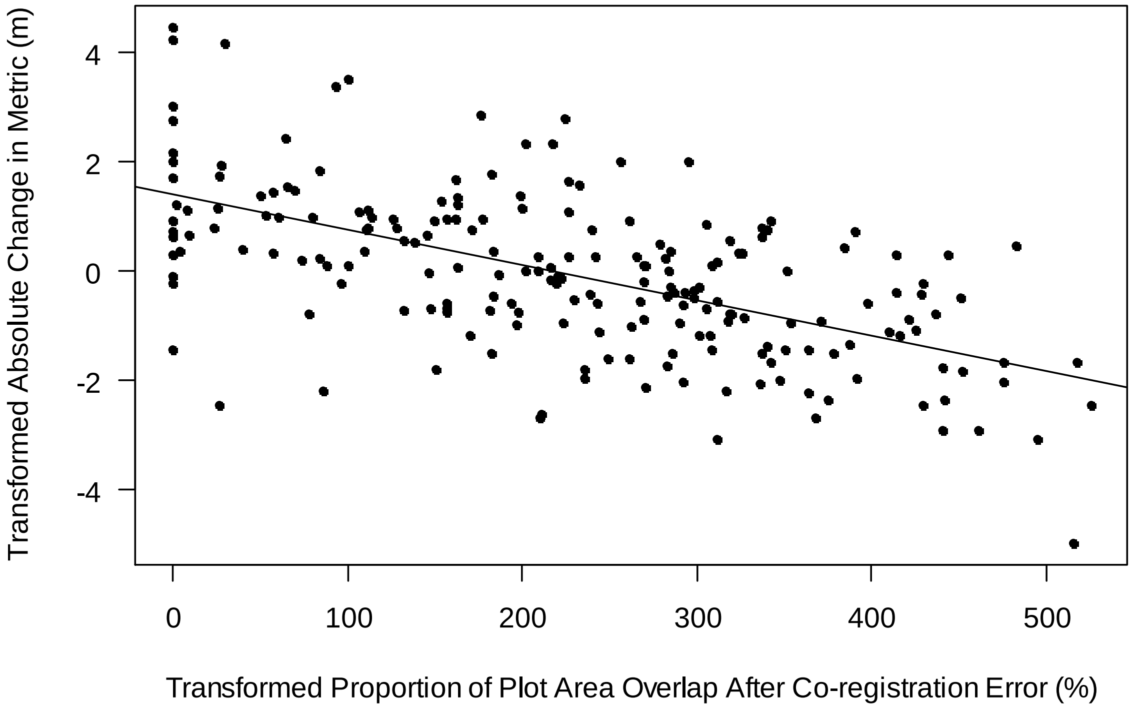

Figure 21: The relationship between the transformed H_{20} LiDAR canopy metric and the transformed plot overlap

The PCR>3 LiDAR Canopy Metric

The relationship between the transformed PCR>3 LiDAR canopy metric and the transformed plot overlap is presented in Figure 22. The regression between these two variables had the model form:

$$PCR>3^{(-0.10/1.10)} = 0.175 - (0.0034 * Overlap^{0.45/1.45})$$

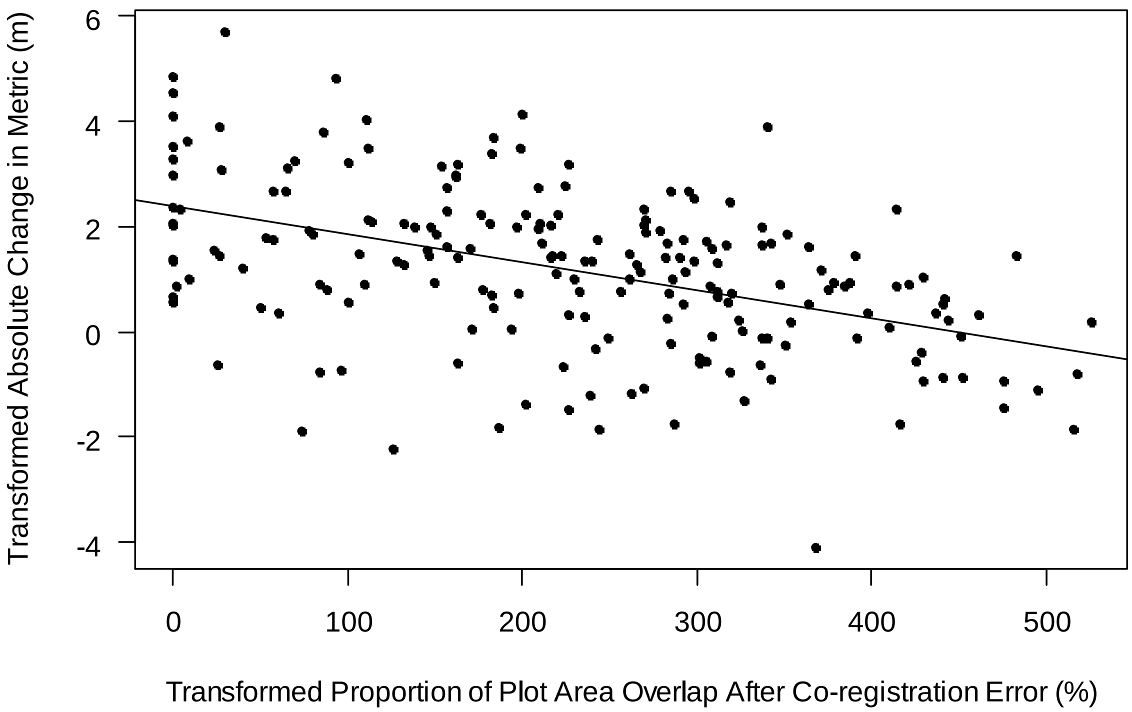

Figure 22: The relationship between the transformed PCR>3 LiDAR canopy metric and the transformed plot overlap

Appendix 4: The Linear Spatial Errors within Each Stratified Group of Plot Co-registration Error:

For comparison, the minimum and maximum linear plot co-registration errors distances for each of the five stratified groups are presented in Table 9. This is compared by the different plot sizes that were sampled.

Table 9: The ranges of linear co-registration error distances within overlap groups by plot size.

Range in Linear Plot Co-regsitration Errors (m) within Each Stratified Group					
	80-100% Overlap	60%-80% Overlap	40%-60% Overlap	20%-40% Overlap	0%-20% Overlap
$400m^2$	0.0 - 3.7	3.8 - 7.2	7.2 - 11.2	11.2 - 15.8	>15.8
$500m^2$	0.0 - 3.9	3.9 - 8.0	6.0 - 12.5	12.5 - 17.5	>17.5
$600m^2$	0.0 - 4.1	4.1 - 8.7	8.7 - 13.7	13.7 - 19.2	>19.2